(51) International Patent Classification⁷: G01N

(21) International Application Number: PCT/US01/31037

(22) International Filing Date: 4 October 2001 (04.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/237,383   4 October 2000 (04.10.2000)   US

(71) Applicant *(for all designated States except US)*: CELADON LABORATORIES, INC. [US/US]; 387 Technology Drive, College Park, MD 20742 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: PETERSON, Raymond, J. [US/US]; 7940 Vanity Fair Drive, Greenbelt, MD 20770 (US).

(74) Agents: GOLLIN, Michael, A. et al.; Venable, Baetjer, Howard & Civiletti, LLP, Suite 1000, 1201 New York Avenue, NW, P.O. Box 34385, Washington, DC 20043-9998 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(54) Title: COMPUTER SYSTEM FOR DESIGNING OLIGONUCLEOTIDES USED IN BIOCHEMICAL METHODS
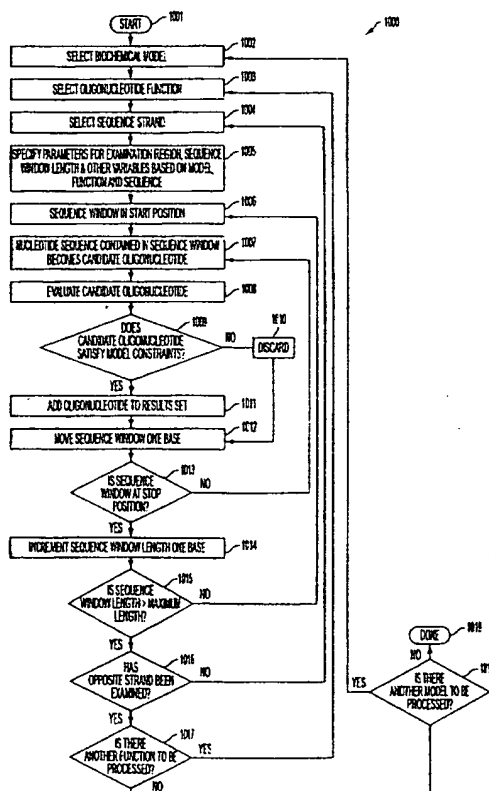
(57) Abstract: A method for determining one or a set of optimal reagent oligonucleotide sequences used in a biochemical method for evaluating a target nucleic acid sequence having a target feature by defining a set of exclusion values or ranking values specific to the biochemical method, generating candidate reagent oligonucleotide sequences, evaluating the candidate reagent oligonucleotide sequences against the exclusion or ranking parameters, selecting at least one optimal reagent oligonucleotide sequence for the selected biochemical method as applied to the target nucleic acid sequence.

WO 02/29379 A2

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

COMPUTER SYSTEM FOR DESIGNING OLIGONUCLEOTIDES USED IN BIOCHEMICAL METHODS

BACKGROUND OF THE INVENTION

The invention is in the field of bioinformatics and provides methods and systems for generating optimal reagent oligonucleotides for use in biochemical methods, for comparing and evaluating biological sequences, for providing sequences of biological molecules in a relational format allowing retrieval in a client-server environment, and for creating libraries of DNA hybridization probes.

The invention provides and methods for comparing and evaluating biological sequences, for providing sequences of biological molecules in a relational format allowing retrieval in a client-server environment, and libraries of DNA hybridization probes.

All populations of organisms exhibit genetic diversity. In any particular population, the extent, kind and structure of genetic diversity is influenced by the biological processes of mutation and recombination, as well as the population genetic processes of natural selection and random genetic drift. The effect of these processes depends on population size, subdivision and history, as well as mating patterns. A newly arisen variant may confer an evolutionary advantage or disadvantage, or it may be neutral. Natural selection may remove a disadvantageous variant from a population, drive a favored variation to fixation, or maintain polymorphism due to balancing effects. Loss, fixation or polymorphism of neutral variations may occur due to chance events. (Hartl, D., and Clark, A., *Principles of Population Genetics, 2nd Ed.*, Sinauer Assocs, Inc., Sunderland, MA. © 1989).

Hybridization methods to score genetic diversity have not realized their potential. A primary cause for this is that software has been unavailable to comprehensively analyze the nucleic acid sequence context of a targeted variation. Because of this, there are differential success rates across laboratories. Laboratories that happen to have researchers who are either lucky or who develop a touch for a method are able to achieve allelic discrimination in some 70 to 90 out of 100 designed assays based on "brute force" approaches alone. Other labs, with less experienced or less lucky researchers, often have little or no success, failing to get even a single assay to perform well. Assessing millions of genetic polymorphisms in tens, hundreds, and thousands of biological samples represents an enormous task.

-1-

In order to more efficiently score genetic polymorphism, a number of molecular biology methods have been developed. One method is "single base extension", a form of nucleotide sequencing. In this method, an oligonucleotide sequencing primer is extended by just one base, and this base is complementary to the targeted variation.

Additional methods include hybridization methods such as oligonucleotide arrays, for example PCT Application WO 99/05324, molecular beacons, Invader, the 5' nuclease method, and DASH (Howell *et al.*, (1999) *Nat. Biotech.* *17*:87-88). The principle underlying these methods is that an oligonucleotide will bind more strongly to a target DNA sequence when there is perfect, complementary Watson-Crick base pairing compared to when there is one or more mismatches between the oligonucleotide probe and the complementary target sequence. Ideally, probe hybridization should be digital. That is, a probe should always hybridize to its perfectly complementary sequence and never hybridize to sequence that is not perfectly complementary.

Despite the recent completion of drafts of the human genome and other genomes, and the identification of millions of genetic polymorphisms, to date only a tiny fraction of genetic diversity has been studied with respect to medically and commercially important traits. The small number of studied polymorphisms is largely due to the large amount of work required of conventional laboratory methods and processes. One aspect of conventional methods that is particularly labor intensive is the design of assays, and most particularly the design of oligonucleotide primers or probes used therein. Present design methods often results in sub-optimal assays that require extensive laboratory optimization in order to obtain meaningful signals while keeping nonspecific biological background interference, primer dimerazation, and oligonucleotide secondary structure formation to a minimum (Saiki, *et al.* (1985) *Science 37*:170-172). This is especially true for methods that use hybridization probes to discriminate among genetic variations.

It would be highly useful to apply SNP scoring and especially hybridization methods to the study of genetic diversity on a large scale. For example, it would be useful to study the association between certain variations and susceptibility or resistance to specific diseases, or to drug response. To accomplish these benefits will require the large-scale design of oligonucleotides to be used as PCR primers, allele-specific hybridization probes, and to perform other functions. Further, it will require storing a vast amount of data in such a way as to ease later querying and retrieval. What is needed is an improved process and

5   methods suitable for large-scale design of genetic diversity assays and systems and methods for organizing large amounts of data used in genetic diversity studies.

The first previous approach is PrimerExpress™ Software from Applied Biosystems. This software functions as a calculator where the user must input each sequence individually. Thus, comprehensive examinations are not performed. This software does not

10   allow specification of the targeted genetic variation, does not automatically examine both the forward and reverse strands of the DNA molecule, does not automatically evaluate primer and probe sequences for more than one model, and does not communicate with a central database. Better software would be process oriented, such that it leads the user through the design process, requiring little user interference. Such software would also

15   operate in batch mode, being able to process a queue of variations.

The second previous approach is MeltCalc software. This is implemented in an Excel spreadsheet. This software functions as a calculator, but also examines some of the surrounding sequence. It appears that this software (PrimeExpress™) does not perform a comprehensive examination, does not automatically examine both the sense and antisense

20   strands of the DNA molecule, does not evaluate primers in addition to probes, is specific to one model, does not communicate with a central database and is not process oriented.

Many molecular biology methods for scoring genetic variation require the use of one or more reagent oligonucleotides. Each of these reagent oligonucleotides performs a separate function and these functions are well known in the art. These functions include, but

25   are not limited to, forward PCR primer, reverse PCR primer, sequencing primer, allele-specific hybridization probe, anchor probe, invader probe, and reporter-probe. Typically, many candidate oligonucleotides can be considered for each function. The problem is to choose typically one oligonucleotide for each function such that the oligonucleotides for all functions perform well in combination to produce excellent allelic discrimination. In

30   addition it is important to design reagent oligonucleotides that are not cross reactive or inhibitory; for example, to minimize primer dimerization or reagent oligonucleotide cross complementarity, so that the biochemical method employed to evaluate target nucleic acid sequences is most efficient.

Prior approaches resulted in sub-optimal assays because only a few of the candidate

35   reagent oligonucleotides were examined one at a time by researchers. This is slow,

-3-

5   laborious and resulted in many failed assays and much laboratory time and cost to optimize
     reaction conditions.

                              SUMMARY OF THE INVENTION

             The invention provides computer software products to be used by scientists.
10   Compared to existing software, the provided computerized method more quickly designs
      better reagent oligonucleotides for performing biochemical methods.

             The invention automatically performs a comprehensive examination of nucleotide
      sequences adjacent to target features within target nucleic acid sequences such that all
      candidate reagent oligonucleotide sequences (e.g., primers and probes) are available for
15   evaluation by one or more defined biochemical methods. The sequence to be examined
      optionally originates from a database, and the assay design selected by the model or chosen
      by the user is optionally committed to a database. The comprehensive examination
      algorithm is distinct from the model that is applied to evaluate potential primers and probes.

             According to one aspect of the invention, a computer system is used to methodically
20   and comprehensively examine the nucleotide sequence flanking a targeted polymorphism.
      This examination is comprehensive in that, for a given method, all candidate reagent
      oligonucleotides that can be used in the method to query the targeted feature (e.g., single
      nucleotide polymorphism) are considered. One or more biochemical models (e.g.,
      Polymerase   Chain   Reaction,   Reverse   Transcriptase-Polymerase   Chain,
25   ReactionNucleotide™ Sequencing, Fluorescent in situ hybridization, allele-specific
      oligonucleotide hybridization (ASOH), dynamic allele-specific hybridization (DASH),
      antisense oligonucleotide chemistry, nucleic acid hybrid chemistry; DNA/RNA repair, etc.)
      can be applied to a targeted feature. Each model may have a unique set of possible
      oligonucleotides. For a given model, each possible oligonucleotide is evaluated with respect
30   to the variables, parameters, and constraints of the model. An oligonucleotide is retained for
      further analysis only if it satisfies the exclusion constraints of the model. The resulting list
      of oligonucleotides may have zero or more entries. These entries are sorted based on one or
      more ranking parameters of the model.

             The steps involved are: inputting nucleotide sequence which contains a targeted
35   polymorphism; examining one strand (sense or antisense) for one allele of the polymorphism
      so as to determine an oligonucleotide that perfectly matches the complement of the allele;

                                           -4-

5 comparing this oligonucleotide to the complement of alternate alleles in order to assess the
ability of the oligonucleotide to provide predicted allelic discrimination; saving the
oligonucleotide for further analysis only if it satisfies the constraints of the model; repeating
until all possible oligonucleotides have been considered; optionally repeating for all alleles;
optionally repeating for the opposite strand; optionally repeating for all specified models;
10 sorting the resulting lists based on one or more model variables; optionally presenting the
lists to the user for further evaluation and selection; optionally choosing the oligonucleotides
that best satisfy predetermined model constraints; optionally repeating until all targeted
polymorphisms have been processed.

In sum, the invention provides a method for determining an optimal reagent
15 oligonucleotide sequence for use in a biochemical method for evaluating a target nucleic
acid sequence having a target feature, the method comprising the steps of: defining a set of
exclusion values and/or ranking values specific to the biochemical method, defining a
sequence window adjacent to the target feature, generating candidate reagent oligonucleotide
sequences complementary to one or both of the sense and antisense strands of the target
20 nucleic acid sequence within the sequence window, the reagent oligonucleotide sequences
having a length less than or equal to the sequence window, evaluating the candidate reagent
oligonucleotide sequences against the exclusion and/or ranking parameters, selecting at least
one optimal reagent oligonucleotide sequence for the selected biochemical method as
applied to the target nucleic acid sequence.

25 The invention also provides a computer readable data storage medium storing a
computer readable program code means for causing a computer to perform the steps of the
provided method and a computer system comprising the data storage medium.

Additionally, the invention provides a process of manufacturing reagent
oligonucleotides comprising using reagent oligonucleotide sequences from the provided
30 computer system in a nucleic acid synthesizer to produce the selected reagent
oligonucleotides.

Further, the invention provides a kit of a predetermined number of reagent
oligonucleotides optimized for a biochemical method used in evaluating a target nucleic acid
sequence, the reagent oligonucleotides made by the provided process.

35 Lastly, the invention provides a method of ordering a kit of reagent oligonucleotides
by accessing the provided computer system, inputting the desired target nucleic acid

-5-

5   sequence, generating a set of reagent oligonucleotide sequences useful in one or more biochemical methods used in evaluating said target nucleic acid sequence, selecting and ordering the desired sequences from the list of generated reagent oligonucleotide sequences, synthesizing a kit of reagent oligonucleotides based on the selected reagent oligonucleotide sequences, and shipping said kit of reagent oligonucleotides.

10

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows several recognized nucleic acid sequence formats. Upper and lower case used for emphasis only. Brackets ([]) and slash (/) can be any delineating characters.

15   Sequences shown in the 5' to 3' direction. "A" shows sequence in WhiteHead-Affymetrix format, "B" shows sequence in IUPAC format, "C" shows sequence in allelic format, and "D" shows sequence as it appears in a database (*e.g.*, BLAST or FASTA formats) where the number is the position in specified sequence and the letter designation is a variant allele.

Figure 2 shows the numbering systems used to indicate the relative positions of

20   nucleotide bases used in the present invention. "A" shows a portion of a sequence window region for an allele-specific hybridization probe where the position of the polymorphism is designated "0" and the nucleotides 5' or upstream to the polymorphism are descending negative and 3' or downstream are ascending positive. "B" shows an oligonucleotide designed by the present invention that could be used as an oligonucleotide-probe, or primer;

25   the numbering scheme herein deisgnates the first nucleotide of the oligonucleotide as "1" and 5' or upstream nucleotides are number in ascending numerical order.

Figure 3 is a depiction of a nucleotide sequence under examination having an Examination Region, "A", Sequence Window, "B" and Open Positions "C".

Figure 4 shows an example of the complementary candidate oligonucleotide

30   generated using the present invention in a sequence window. The illustrative target nucleic acid sequence of 15 base pairs contains polymorphic base in the middle third of the oligonucleotide. "A" shows a sense strand of the target nucleic acid and the candidate reagent oligonucleotide that is complementary thereto at polymorphic base C; "B" shows a sense strand of the allele with polymorphic base T; "C" shows an antisense strand of the

35   allele with polymorphic base G; and "D" shows an antisense strand of the allele with polymorphic base A.

5   Figures 5A-O are screen shots of the graphical user interface screens (desktop version) provided for accepting user queries, inputs of nucleic acid sequence and for the display of results generated using the present invention.

Figure 5A shows the screen shot depicting a scatter plot of results obtained from the invention for adenine $\rightarrow$ guanine purine to purine point mutation in the *drd2* encoding
10   nucleic acid sequence. The abscissa measures the fluorescence emitted by a hybridization probe to the allele with an A polymorphism, and the ordinate measures the fluorescence emitted by a hybridization probe to the allele with an G polymorphism, each dot represents a sampled organism.

Figure 5B shows a plurality of genetic sequences available in the database for a
15   project and the highlighted selection of the IGF1 encoding nucleic acid sequence.

Figure 5C shows the genetic variations entered into the database for the IGF1 gene and the selection of the IGF1_P325L allele, wherein the polymorphism changes amino acid 325 of the encoded protein. The targeted polymorphism and flanking sequence are examined with respect to the correct sequence format.

20   Figure 5D shows the parameter selection screen for the probe model with default values. Default parameter values that rarely need to be adjusted are shown. On the left of the screen shot of Figure 5D are the variables of the biochemical model for which the user can enter parameter values. On the right, the probe constraints box contains input boxes for W_{min}, W_{max}, MinTm and MaxTm. The chemistries box has check boxes for selecting one of
25   two biochemical models. The choices shown are standard phosphoramadite chemistry and PropyneT chemistry (see Table 1). The difference between the two models are the values used to calculate T_m.

Figure 5E shows a results screen where the saved (within exclusion parameter constraint values) versus discarded (outside exclusion parameter constrain values) reagent
30   oligonucleotide sequences are sorted by ΔT_m for the standard chemistry of the sense strand.

Figure 5F shows a results screen where the saved (within exclusion parameter constraint values) versus discarded (outside exclusion parameter constrain values) reagent oligonucleotide sequences are sorted by ΔT_m for the standard chemistry of the antisense strand.

35   Figure 5G shows a results screen wherein the determined oligonucleotides are sorted by ΔT_m for the PropyneT chemistry of the sense strand.

Figure 5H shows a results screen wherein the determined oligonucleotides are sorted by ΔT_m for the PropyneT chemistry of the antisense strand.

Figure 5I shows Probe results screen with oligonucleotide reporter and quencher defaults.

Figure 5J shows a screen for parameter selection and default values.

Figure 5K shows the results screen for forward PCR primer with primes tiling to the left away from targeted variation.

Figure 5L shows the results screen for reverse PCR primer with primers tiling to the right away from targeted variation.

Figure 5M shows the results screen for primers and showing primers and probes on inputted sequence.

Figure 5N shows a screen wherein the assay is given an arbitrary name, the genetic locus or nucleic acid sequence under examination, the name of the method used and variable fields that identify the submitter of the assay.

Figure 5O shows the selection screen for a Sequence Tagged Site name and the name for both forward and reverse oligonucleotide primers.

Figures 6A-6F show method specific forward and reverse primer annealing sites and allele specific probe sites relative to a polymorphic site. In all of these figure the areas indicated by shading show the region of the target nucleic acid where a reagent oligonucleotide is complimentary and would optimally bind as a primer or probe.

Figure 6A shows a nucleotide fragment with the location of Examination Regions for the 5' nuclease method (TaqMan™, Applied Biosystems, Inc.). "A" shows the site at which the forward PCR primer anneals, "B" shows the site at which the allele specific probe(s) anneal(s), "C" shows the site at which the reverse PCR primer anneals, "X" indicates the targeted feature.

Figure 6B shows a nucleotide fragment wherein the location of Examination Regions for 5' nuclease method with minor groove binding (MGB) probes (TaqMan, Applied Biosystems). "A" shows the site at which the forward PCR primer anneals, "B" shows the site at which the allele specific probes anneal with conjugated a minor groove binder group (601), "C" shows the site at which the reverse PCR primer anneals, "X" indicates targeted feature.

5   Figure 6C shows a nucleotide fragment with the location of Examination Regions for the anchor method (Light Cycler, Roche). "A" shows the site at which the forward PCR primer anneals, "B" shows the site at which the allele-specific probe(s) anneal, "C" shows the site at which the anchor probe anneals, and "D" shows the site at which the reverse PCR primer anneals. "X" indicates the targeted feature.

10   Figure 6D shows a nucleotide fragment with the location of Examination Regions for the Invader method (Third Wave). "A" shows the site at which the allele-specific Invader Probe anneals, "B" shows the site at which the signal probe with line indicating displacement of 5' end anneals. "X" indicates the targeted feature.

        Figure 6E shows a nucleotide fragment with the location of Examination Regions for
15   the single base extension method (Orchid Bioscience, Inc.) "A" shows the site at which the forward PCR primer anneals, "B" shows the site at which the sequencing primer anneals, "C" shows the site at which the reverse PCR primer anneals. "X" indicates the targeted feature.

        Figure 6F shows a nucleotide fragment with the location of Examination Regions for
20   the DASH method. "A" shows the site at which the forward PCR primer anneals, "B" shows the site at which the allele specific probe(s) anneal, "C" shows the site at which the reverse PCR primer anneals. "X" indicates the targeted feature.

        Figures 7P-7V are web-based client screen shots for a system analogues to the desktop version of the software package as in the present invention as depicted in Figures
25   5A-5O.

        Figure 8 shows a database of a relational database containing information such as target nucleic acid sequence data and allelic sequences thereof, target nucleic acid sequences desired as targets for hybridization or cloning, reagent oligonucleotide sequence information, the authentication information of the scientist; investigator; or user submitting
30   or retrieving information, biochemical methods and their exclusion parameters and constraints values and ranges thereof.

        Figure 9 is a flow chart showing an overview of the process of the present invention.

        Figure 10 is a flow chart showing the comprehensive evaluation and examination algorithm of the present invention.

5   Figure 11 shows the diagram for ProbeITy™ (Celadon) software package which incorporates the present invention which integrates the algorithm and evaluation process of Figure 10.

10   <u>DETAILED DESCRIPTION</u>

In describing preferred embodiments of the present invention illustrated in the drawings, specific terminology is employed for the sake of clarity. However, the invention is limited to the specific terminology so selected. It is to be understood that each specific element includes all technical equivalents which operate in a similar manner to accomplish a

15   similar purpose. Each reference cited here is incorporated by reference in its entirety as if each were individually incorporated by reference.

Disclosed is an improved computer-aided process and method for generating and evaluating oligonucleotides used in a variety of customized or commercially available molecular biology methods. The present invention is primarily illustrated, but is not limited

20   to, the design of optimized oligonucleotide primer pairs for the 5' nuclease method (TaqMan™, Applied Biosystems, Inc); this results in allelic discrimination with a success rate above 90%.

The invention also provides a relational database system for storing, organizing, and displaying target nucleic acid and reagent oligonucleotide sequence information together

25   with sequence annotations; such as allele variation detail, the information important to evaluating, studying, and cataloging the genetic diversity of a species.

The term "adjacent" as used herein describes the distance of the sequence window from the target feature (in either the 3' or 5' directions) ranging in practice from overlapping or encompassing to a distance of 30 kb away from the target feature, the distance being the

30   limits of the amplification technology to "reach" the target feature from that distance. For sequence windows that do not overlap the target feature these sequence windows are primarily used for selecting reagent oligonucleotide sequences that act as amplification primers. It is known in the art that long range PCR amplification methods can generate amplicons that are more than about 40kb in length (Barnes *et al.* (1994) *Proc. Natl. Acad.*

35   *Sci. USA 91:* 2216-2220, see also http://www.genecraft.de/products/48.htm).

The term "biochemical method" or "biochemical model," and the like, as used herein shall include any chemical and biochemical protocols, methods, models, reactions,

5   experiments, diagnostic, techniques, and assays for evaluating target nucleic acid sequences. The term "evaluating" as used herein includes all methods of nucleotide processing such as amplification, reverse transcription, transcription, translation, immobilization, cloning, hybridization, sequencing, antisense biochemistry, RFLP, DNA or RNA repair, mutagenesis biochemistry, site-specific mutagenesis, and mutagenic recombination.

10   The terms "optimal," "optimized" and the like as used herein describes the best fit of a reagent oligonucleotide sequence within the exclusion and ranking parameters of a selected biochemical method. An "optimal set" or "solution set" of reagent oligonucleotides also satisfies compatibility criteria.

         An "evaluating parameter" refers to physical or chemical properties of the candidate
15   reagent oligonucleotide sequences necessary for the sequences to perform well in a given biochemical method. Examples include thermodynamic parameters, amplicon parameters, oligo parameters, secondary structure parameters, and sequence parameters, all of which are explained in greater detail below. For each evaluating parameter, there are exclusion, ranking, and/or compatability values, which are predetermined constraint values for a
20   particular parameter. Exclusion values (typically a minimum and a maximum or presence and absence) are used to test candidate sequences, in a pass/fail mode. For ranking values, the constraints provide an ideal value or range of values against which candidate sequences are evaluated. For compatibility values, the constraints ensure that candidate reagent oligonucleotides of a candidate set are compatible, e.g. not self inhibitory, cross reactive, or
25   cross complementary with each other (e.g., formation of secondary structures with each other, primer dimerization, competitive inhibition, accidental ligation, circularization, or inherent catalytic activity such as ribozymes or lariats).

         The term "reagent oligonucleotide sequences" as used herein generally refers to the sequence information (e.g., order of adenines, guanines, cytosines, and thymidine) for a
30   oligonucleotide used in as a reagent in a Biochemical method. This information can be in written form, computer readable form; such as but not limited to text, ASCII, RTF, MSWord, or any other computerized text format, or can be in a computer readable form that has been encrypted, formatted in HTML, or in any other coded form. The term "reagent oligonucleotides" is generally used here to mean the physical oligonucleotide which has
35   been synthesized using any oligonucleotide synthesis technique known in the art which is

5   used as a reagent in a biochemical method. A person of ordinary skill will understand that reference to a sequence connotes the oligonucleotide, and vice versa.

The term "sequence window" means a section of either or both of the sense or antisense sequence of the target nucleic acid sequence selected for generating potential candidate reagent oligonucleotide sequences that are complementary to the target nucleic

10   acid sequence (in particular to evaluate the target feature). The sequence window has a length of nucleotide bases/base pairs that is less than or equal to the length of the entire target nucleic acid sequence, and is greater than or equal to the minimum useful length of a candidate oligonucleotide. In the typical embodiment where the sequence window length is less than the entire target nucleic acid, all possible candidate reagent oligonucleotides

15   complementary to the target nucleic acid within the sequence window are considered. The possible candidate oligonucleotides to be considered are generally constrained by the exclusion values for the size parameter. For example if the maximum length of a reagent oligonucleotide sequence (*e.g.*, for a primer or probe), as constrained by the length parameter, is 35 bp then the sequence window is 35 bp. If the length parameter constraint

20   minimum is 10 bp, then all possible reagent oligonucleotide sequences of lengths 10-35 within the sequence window will be considered. The location of the sequence window and adjacency to the target feature in the target nucleic acid sequence, is dependent on the particular biochemical method used and type of reagent oligonucleotide required. According to the invention, the sequence window can be repositioned along the target nucleic acid

25   sequence according to the constraints of the biochemical method. For example, if the biochemical method employed is PCR and the reagent oligonucleotide sequences evaluated are forward and reverse primers; the PCR method able only to generate amplicons having a maximum length of 100 bp and use primers of maximum length 30 bp, then a 30 bp sequence window will be created and repositioned along the desired amplicon containing the

30   target feature until all candidate reagent oligonucleotides sequences useful in making such an amplicon are generated for consideration. The sum of the range of positions within which the sequence window may be stepped (or incremented or repositioned) can be referred to as the "examination region," which in this case would be 100 bp (see Figure 3). For a probe, the sequence window and the examination region would typically be coextensive (the

35   same).

-12-

5   The term "target nucleic acid sequence" means a double or single stranded DNA or RNA, chimera/hybrid, or analogue comprising a sequence of adenines, guanines, cytosines, thymidines, or uracils, under evaluation. The target nucleic acid sequence has one or more target features. "Target feature" means a sub-sequence of a target nucleic acid sequence of interest to the researcher, such as a single nucleotide polymorphism, a multimeric

10   subsequence of the target nucleic acid sequence, a cloning sequence, the entire target nucleic acid sequence under examination, a codon, an exon, an intron, a telomere, a viral sequence, a transposon, a noncoding region, a promoter, an enhancer sequence, an expressed sequence tag, and a sequence tagged site.

15   *Oligonucleotide Design*

        A key aspect of the invention is the process of methodically performing a comprehensive examination and evaluation of the nucleotide sequence flanking a polymorphism so as to arrive at a combination or set of oligonucleotides per biochemical method or model. This set is optimal in that it maximizes the predicted chance of yielding

20   the highest allelic discrimination. The present invention identifies all oligonucleotides that satisfy the constraints of a specified biochemical model.

        An overview of the process in Figure 9 (900), for each reagent oligonucleotide sequence, to generate all possible candidate reagent oligonucleotide sequences that have a

25   length between some non-constraining minimum and some non-constraining maximum number of nucleotide positions. Each of these candidate oligonucleotides is evaluated with respect to some biochemical model parameterized according to the molecular biology method and oligonucleotide function under consideration. The candidates that satisfy the constraints of the model are retained and sorted based on one or more variables of the

30   model. The zero or more candidates that maximize the predicted chance of producing excellent allelic discrimination are at, or near, the top of the sorted list.

        First an allele of a nucleic acid sequence having at least one polymorphic site is entered via a graphical user interface (901) and read (902). Optionally, the data comprising the entered nucleotide sequence can be saved in a memory buffer for later access and to

35   facilitate the creation of a data entry in a relational database. The software then checks whether the entered nucleic acid sequence conforms to known sequence file formats (903)

5   which include but are not limited to, FASTA, MSF, NBRF, NEXUS, PHYLIP, raw ASCII,

text, WhiteHead-Affymetrix format, or IUPAC formats (Baxevanis *et al.*, eds.,

Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Wiley-

Interscience, New York, © 1998, pgs. 359-362 (Appendix 2)). The software is customizable

to default to one biochemical method or can be modified to include a plurality of user

10   selectable (905, 906) biochemical methods that employ oligonucleotides (904a, 904b).

Upon entry of the allelic nucleic acid sequence under study, the software runs the algorithm

(907) based on the selection parameters required by the respective biochemical method and

determines a set of putative oligonucleotides that can be used in the respective biochemical

method to study the allelic nucleic acid sequence. Concurrently, the software of the

15   underlying method of the present invention performs an optimization subroutine which

selects (908) among the several determined oligonucleotide sequences for the optimum for

that biochemical method. The selected oligonucleotides are the candidate set and are

displayed in descending optimal value (910). The scientist/user then selects which

oligonucleotides she wishes to employ (911) and can then restart the process for a

20   oligonucleotide selection in another biochemical method (912) and for any additional

polymorphic regions of the entered allelic nucleic acid sequence that may require another

run of the software (913, 914). After all of the oligonucleotides corresponding to the entered

nucleic acid sequence and desired biochemical method(s) have been determined and

selected, the user may optionally save the information into a relational database (915) that

25   serves to catalogue the genetic variation of the entered nucleic acid sequence (916) and the

determined oligonucleotides useful in its study. The user may opt to enter another allelic

sequence, another polymorphism for an existing sequence, or terminate the program (917).

Algorithm and Comprehensive Examination

30   The invention defines discrete steps of the oligonucleotide assay design process and

leads users through each step in a wizard-like fashion. The invention includes means for

inputting target sequences into the software, validating target sequences to ensure proper

formatting, displaying and describing the entered target nucleic acid sequence for user

verification and comfort, defining a sequence window for each functional reagent

35   oligonucleotide sequence in the biochemical method from which candidate reagent

-14-

oligonucleotide sequences are generated, enabling a user to choose a predefined set of parameter values, enabling the user to customize parameter values for a specific design run instance, automated generation, constraint evaluation and scoring of candidate reagent oligonucleotide sequences for each window in a method-specific manner, displaying and enabling a computer graphical user interface ability to select one or more candidate oligonucleotides; or sets of candidate reagent oligonucleotides for consideration at subsequent steps of the design process, enabling the navigation back and forth through the design process so as to modify previous actions, enabling the selection of one or more biochemical methods, or individual oligonucleotides of an assay, enabling the fully automated design process by having the system select entire assays rather than individual candidate oligonucleotides, enabling the processing of more than one sequence either by stepping through the design process for each sequence consecutively or to automatically generate assays for each sequence in batch mode.

Further, the assay design process is integrated into an E-commerce platform that includes, user name and password protection for system entry and security, user registration with contact information, marketing survey at time of registration, user accounts that record personal contact, billing and shipping information and that track order history, shopping cart features that allow users to purchase one or more assays per sequence. The system is also designed to feed directly into a manufacturing operation. Specifically, the purchased oligos and relevant manufacturing instructions can be electronically downloaded automatically or manually and imported into the instruments that synthesize the actual oligonucleotides.

One novel aspect of the invention is the definition of a separate sequence window for each reagent oligonucleotide of an assay. Once each sequence window is chosen, the process of generating, evaluating and scoring candidate oligonucleotides is the same.

The generation of the list of candidate reagent oligonucleotides is methodical. The steps are:

- Start at the first sequence position of the window. the first candidate oligonucleotide starts at that position and has length determined by the parameter MinOligoLength or the end of the window, whichever is shorter.

- Continue to generate oligonucleotides that start at this sequence position by incrementing the length by one until attainment of either the MaxOligoLength or the end of the window.

-15-

5   - Move to the next sequence position and repeat the process until the start sequence position is less than MinOligoLength positions from the end of the sequence window.

The evaluation of the min and max constraints is also methodical. For each parameter, the corresponding property is calculated for each candidate reagent

10   oligonucleotide sequence. Each property is evaluated with respect to the min and max values. If the property is greater than or equal to the min and less than or equal to the max, the candidate reagent oligonucleotide sequence satisfies the constraints and is marked as pass. Otherwise, the candidate reagent oligonucleotide sequence fails and is marked as fail.

Oligonucleotides that pass the constraints are then scored according to the ranking

15   parameter. For each parameter, this function measures the distance of the oligonucleotide property from the ideal value and then multiplies this distance by the Weight parameter. Scores for each parameter are then summed to arrive at a single, composite score. Distance is measured by the absolute value of the difference between the ideal value and the property value. However, other distance measures are equally viable, such as the squared difference

20   or the unit variance of the property based on a statistical or empirical distribution. Min and max constraints can be applied to this score. Typically, candidate oligonucleotides are displayed to the user sorted by score.

Alternatively, candidate oligonucleotides are not evaluated for the min and max constraints, but are simply scored and sorted on this score. Another option is to score

25   candidate oligonucleotides and then to apply the min and max constraints. One set of min and max constraints could be applied to the score property.

The parameters used in the present invention are described below. Each of these parameters typically has four sub-parameters: Min, Ideal, Max and Weight. The Min and Max values are constraints that a candidate oligonucleotide must satisfy; to be considered in

30   an assay, a reagent oligonucleotide sequence's property must be greater than or equal to the min and less than or equal to the max. The Ideal parameter is the value most desired by the user; this value is used in the scoring function that ranks candidate oligonucleotides that satisfy the min and max constraints. The Weight parameter is used in the scoring function and allows a user to adjust the relative importance of each parameter.

35   Exemplary evaluating parameters and constraint values used in the invention include some or all of the following, depending on the biochemical model and other factors. A

-16-

5   person having ordinary skill in the art will readily be able to select from these parameters
     and values and others.

Thermodynamic Parameters

10   OligoTm

     The temperature at which half of the oligos are in the single stranded state and half are in the
     double stranded state.   Melting temperature is critical to oligonucleotide performance
     because genomic applications are tuned to specific Tms. The T_m of PCR primers is typically
     58-60 degrees. Other genomic applications, such as TaqMan, require hybridization probes

15   that have a target T_m of 68-70 degrees.

     Typically $T_m = \dfrac{dH}{\left(dS + R\ln\left(\frac{c}{4}\right)\right)} - 273.15$,

     wherein H is enthalpy, S is entropy, R is the gas constant (1.987 cal/mol·K), and C is
     concentration in units Molar.

20   BufferMg++

     The concentration of Mg++ divalent ions in the buffer. Mg++ is a necessary cofactor for
     DNA-directed DNA polymerases. Higher concentrations increase T_m. Most biochemical
     methods utilize values ranging from 2.0 mM to 5.0 mM.

25   BufferK+

     The concentration of salts such as K+ and Na+ help enhance hybridization by neutralizing
     the negative charge of the backbone. Higher salt concentrations increase T_m. Typical values
     range from 40 mM to 60 mM.

30   DivalentMultiplier

     The relative affect on T_m of divalent ions, such as Mg++, as compared to monovalent ions
     such as K+, is a matter of empirical uncertainty. Some researchers report a 100-fold
     difference, while others report a 150 fold difference. The DivalentMultiplier parameter
     allows this difference to be adjusted by the user.

35

     Oligonucleotide Concentration

5   Increasing the concentration of an oligonucleotide increases its T_m. Typical values range

from 0.1 μM to 1.0 μM.

AmpliconTm

An amplicon is the sequence amplified by a left and a right PCR primer. The melting

10   temperature of an amplicon should be high enough so as to not interfere with

oligonucleotide hybridization (55-70) or polymerase extension (70-75), and low enough that

the two strands of the amplicon are fully disassociated at the melting temperature (94-96).

15   **Amplicon Parameters**

Amplicon Length

1. The number of base pairs in the amplicon, including the left and right PCR primers.

PCR is typically most efficient for short amplicons. Typical limits are 10 to 150 base.

20   Other applications, such as sequencing, SSCP or dHPLC, target amplicons within a

certain range, such as 150 to 300 bp.

AmpliconGC

The number of Gs plus the number of Cs in the amplicon divided by the total number of

25   base pairs in the amplicon. Typically, 50% GC/AT content is desired. Uneven GC content

can result in poorly understood thermodynamics. It can also reduce the polymerases' ability

to synthesize a new strand by causing slippage or by the presence of secondary structures

that inhibit the polymerases' ability to traverse the strand.

30   **Oligonucleotide Parameters**

OligoLength

The number of base pairs in the oligonucleotide. In the present invention, OligoLengthMin

and OligoLengthMax are the key parameters used to methodically generate all possible

35   candidate oligos for a sequence. The typical range is between 10 and 40 base pairs. In the

absence of accurate T_m prediction, early PCR rules suggested oligos of about 20 base pairs

-18-

5 and about 50% GC content. In some applications, such as allelic discrimination, shorter hybridization oligos are preferred because they more easily discriminate single base allelic differences.

OligoGC

10 The number of Gs plus the number of Cs in the oligonucleotide divided by the total number of nucleotides in the oligonucleotide. Typically, 10% to about 90%, with 50% GC/AT content preferred. Uneven GC content can result in poorly understood thermodynamics. It can also reduce the polymerases' ability to synthesize a new strand by causing slippage or by the presence of secondary structures that inhibit the polymerases' ability to attach to the

15 double strand formed by the oligonucleotide and its template.

OligoMonoNucRunLength

The longest contiguous stretch of identical nucleotides. Runs of a single nucleotide can cause polymerase slippage or secondary structures that inhibit the ability of the polymerase

20 to attach to the double strand formed by the oligonucleotide and its template.

Oligo5EndLinker

The chemical entity used to link a molecule to the end of an oligonucleotide. This information may be descriptive and used only in the manufacture of the oligonucleotide. In

25 other situations, the effect of the linker is taken into account when predicting T_m.

Oligo5EndModification

The molecule that is linked to the 5' end of the oligonucleotide. There are a large number of modifications of which the most frequent are fluorescent dyes. This information may be

30 purely description and used only in the manufacture of the oligonucleotide. In other situations, the effect of the linker is taken into account when predicting T_m.

Oligo5EndTail

DNA sequence added to the 5' end of an oligonucleotide. This is sometimes used to allow

35 for a second round of PCR using a standard PCR primer; to create secondary structure, such

5   as for molecular beacons; or to increase the length of the primer, such as for single base extension.

Oligo5EndAllowedBases

The bases that are allowed at the 5' end of the oligonucleotide. Often, all four bases
10   (A,G,C,T) are allowed. Other times, such as for hybridization probes, a 5' G is not allowed because the G quenches the signal of the fluorescent dye.

Oligo5EndLeftPosition

The most 5' position in the template sequence at which an oligonucleotide can start. This
15   parameter is useful for precisely locating an oligonucleotide in a template sequence.

Oligo5EndRightPosition

The most 3' position in the template sequence at which an oligonucleotide can start. This parameter is useful for precisely locating an oligonucleotide in a template sequence.
20

Oligo3EndLinker

Analogous to 5' parameter, but at the 3' end of the oligonucleotide.

Oligo3EndModification

25   Analogous to 5' parameter, but at the 3' end of the oligonucleotide.

Oligo3EndTail

Analogous to 5' parameter, but at the 3' end of the oligonucleotide.

30   ### Oligo3EndAllowedBases

Analogous to 5' parameter, but at the 3' end of the oligonucleotide.

Oligo3EndLeftPosition

The most 5' position in the template sequence at which an oligonucleotide can end. This
35   parameter is useful for precisely locating an oligonucleotide in a template sequence.

5   <u>Oligo3EndRightPosition</u>

The most 3' position in the template sequence at which an oligonucleotide can end. This parameter is useful for precisely locating an oligonucleotide in a template sequence.

<u>Oligo3EndAnalysisLength</u>

10   The 3' end of a PCR primer is given special consideration because this is where the polymerase attaches. Poor hybridization or secondary structure at this end of the oligonucleotide may have more impact on performance than in other locations of the oligonucleotide. The typical analysis length is the last 5-7 nucleotides of the oligonucleotide.

15

<u>Oligo3EndGPlusC</u>

The number of Gs and Cs in the last n positions of the oligonucleotide, where n is equal to the Oligo3EndAnalysisLength. Too many Gs and Cs mean the 3' end 'may hybridize in many locations throughout the genome. Conversely, too few Gs and Cs may mean that the

20   3' end of the oligonucleotide is hybridized too weakly for proper polymerase interaction. Often, no more than 3 Gs and Cs are allowed.

<u>Oligo3EndDeltaG</u>

A thermodynamic measure of the hybridization strength of the 3' end, where deltaG is the

25   change in free energy. A measure complementary to Oligo3EndGPlusC that in addition to the number of Gs and Cs in the 3' end also takes into account sequence order.

<u>Oligo3EndGCClampLength</u>

The number of contiguous Gs and Cs that end the oligonucleotide. Some researchers feel

30   that ending the oligonucleotide with 2 or 3 contiguous Gs and Cs helps clamp down the 3' end, thereby enhancing the ability of the polymerase to attach to the oligonucleotide-template complex.

Secondary Structure Parameters

35

<u>OligoAlignMatch</u>

-21-

5   The score given to a Watson-Crick base pairing (C:G or A:T). The value of 1.0 is typical. Watson-Crick base pairings are critical to the performance of an oligonucleotide. The goal is for the oligonucleotide to hybridize to one, and only one, location in the template sequence. The performance of an oligonucleotide is compromised if it has too many Watson-Crick matches with itself (hairpin or self-alignment), with another oligonucleotide

10   (pair alignment), or at another location in the template sequence (mis-priming, non-specific hybridization). While zero matches is the desired hairpin, self-alignment, or mis-priming situation, an oligonucleotide typically performs well even if it has 4 or 5 contiguous matches with itself, with another oligonucleotide, or with another portion of the template sequence.

15   OligoAlignMisMatch

The score given to a non-Watson-Crick base pairing (A:G; C:T; etc.). A value of −1.0 is typical. This parameter penalizes mismatches in a sequence alignment because mismatches disrupt hybridization due to Watson-Crick matches.

20   OligoAlignBulgePenalty

Two adjacent nucleotides in a sequence can form Watson-Crick base pairing with two non-adjacent nucleotides of another sequence. This is possible due to the flexibility of the DNA back bone. The unpaired nucleotides of the one sequence bulge out from the helix formed by the paired bases. This situation is heavily disfavored, and so a penalty of −2.0 is typical.

25

OligoAlignMaxBulgeAllowed

The number of nucleotides allowed between two non-adjacent bases that are paired with adjacent bases. Typically, zero or one bases are allowed. Bulges with more bases are energetically unfavorable.

30

OligoHairPinMinStemLength

A hairpin structure is formed when a single strand sequence (such as an oligonucleotide) folds back onto itself due to Watson-Crick base pairing. The helix formed by the paired nucleotides is often referred to as a stem. Studies indicate that the minimum number of

35   adjacent Watson-Crick pairings needed to form a helix stem is 2.

-22-

5   OligoHairPinMinLoopLength

The loop of a hairpin structure consists of the unpaired bases of the oligonucleotide that lie between the bases of the stem. Studies indicate that the minimum number of nucleotides in a loop is 3. The rigidity of the DNA back bone prohibits the formation of Watson-Crick base pairing for shorter loops.

10

OligoHairPinScore

The highest score of all possible hairpin structures. The score reflects the application of the match, mismatch, bulge, and maxbulge parameters. Given the suggested parameter values above, this score is indicative of the longest stem of all possible hairpin structures.

15   Typically, a score of 5 indicates a hairpin structure that impairs the performance of an oligonucleotide. An oligonucleotide that prefers a hairpin structure will not hybridize with the template sequence. The effect of a hairpin is exacerbated because there are typically billions of copies of the oligonucleotide in solution, but only thousands of copies of template. Thus, an oligonucleotide may be much more likely to assume the hairpin structure

20   than the template sequence.

OligoHairPinDeltaG

Some researchers prefer to measure the hybridization strength of hairpin stems using the change in free energy associated with the Watson-Crick matches of the stem. In addition to

25   number of involved bases, change in free energy takes into account the specific bases and the sequence order of the bases.

OligoHairPinTm

Some researchers prefer to measure the consequences of hairpin stems based on the T_m of

30   the stem. This T_m must be sufficiently lower than the T_m of the oligonucleotide when hybridized to its intended target.

OligoPairAlignScore

Helixes can also form when two sequences have Watson-Crick complementarity. This can

35   happen between two copies of the same oligonucleotide or one copy each of two different

-23-

5   oligos. Again, an oligonucleotide that prefers to bind to itself or another oligonucleotide will not hybridize with the target sequence.

OligoPairAlignDeltaG

Some researchers prefer to measure the hybridization strength of pair alignments using the
10   change in free energy associated with the Watson-Crick matches of the alignment. In addition to number of involved bases, change in free energy takes into account the specific bases and the sequence order of the bases.

OligoPairAlignTm

15   Some researchers prefer to measure the consequences of alignments based on the T_m of the alignment. This T_m must be sufficiently lower than the T_m of the oligonucleotide when hybridized to its intended target.

Oligo3EndPairAlignScore

20   Alignments that involve the 3' ends of both oligos are of special importance because the activity of the polymerase is at the 3' end of the oligonucleotide.

Oligo3EndPairAlignDeltaG

Some researchers prefer to measure the hybridization strength of 3' pair alignments using
25   the change in free energy associated with the Watson-Crick matches of the alignment. In addition to number of involved bases, change in free energy takes into account the specific bases and the sequence order of the bases.

Oligo3EndPairAlignTm

30   Some researchers prefer to measure the consequences of alignments based on the T_m of the alignment. This T_m must be sufficiently lower than the T_m of the oligonucleotide when hybridized to its intended target.

Sequence Parameters

35

SequenceFeature5EndBufferLength

-24-

5   The number of nucleotides that must separate the 5' end of a DNA feature, such as a snp, and the 5' or 3' end of an oligonucleotide. This parameter is useful in precisely locating an oligonucleotide. One purpose is to allow enough sequence between an oligonucleotide and the feature for meaningful molecular analysis. An example is that reliable DNA sequence results typically start some 10 to 20 bases from a sequencing primer.

10

SequenceFeature3EndBufferLength

The number of nucleotides that must separate the 3' end of a DNA feature, such as a snp, and the 5' or 3' end of a PCR primer. This parameter is useful in precisely locating an oligonucleotide. One purpose is to allow enough sequence between oligonucleotide and the

15   feature for meaningful molecular analysis. An example is that reliable DNA sequence results typically start some 10 to 20 bases from a sequencing primer.

SequenceFeaturePosition5End

The 5' position of a template sequence at which a feature, such as an exon or a SNP, starts.

20   This parameter is useful in precisely locating oligos that target the feature (*e.g.*, A, T, C, but not G at the 5' end).

SequenceFeaturePosition3End

The 3' position of a template sequence at which a feature, such as an exon or a SNP, ends.

25   This parameter is useful in precisely locating oligos that target the feature (*e.g.*, T, C, G but not A at the 3' end).

Library SequenceAlignScore

The highest alignment score of an oligonucleotide with a library of sequences. This library

30   is typically a compilation of related genes, either from the same organism or from different organisms. If the goal is to amplify a specific gene, then it is undesirable to have substantial Watson-Crick pairings of an oligonucleotide with other genes. Conversely, if the goal is to target a gene family, it is desirable to have substantial Watson-Crick pairings of an oligonucleotide with several genes.

35

General Methodology

-25-

5   For each entered allelic nucleic acid sequence, an examination region is defined (Figure 3). An examination region is a segment of nucleotide sequence comprising 1 to R nucleotide positions (bases), where R ≤ the total number of bases in the sequence, that defines the sequence that must be comprehensively examined. The biochemical model(s) under consideration explicitly or implicitly govern location, length, and purpose. A

10   Sequence Window is a subsegment of examination region comprising 1 to S nucleotide positions, where s ≤ R. Sequence windows of length s are stepped methodically across the examination region. A sequence window is also defined (Figure 3). The initial number of nucleotide positions in this window is set to some non-constraining minimum number. The window is placed in a start nucleotide position. The nucleotide sequence contained in the

15   sequence window becomes the first candidate oligonucleotide and so on until all possible candidate reagent oligonucleotides have been determined (Figure 3,C).

          Figure 10 is a flow chart showing a comprehensive evaluation and examination algorithm. A candidate reagent oligonucleotide (1003) is evaluated (1000) according to the paramaters (1005) of a model that is defined by the particular biochemical (1002) method

20   and function under consideration. If the candidate oligonucleotide (1007) satisfies the constraints of the model (1009), the oligonucleotide is saved for further analysis (1008, 1011). If the candidate fails to satisfy the model, it is discarded (1010).

          Next, the sequence window is moved one base position (1012). The window now defines the nucleotide sequence of the second candidate oligonucleotide. This

25   oligonucleotide is evaluated and saved or discarded. This process of evaluation and stepping is continued until the window reaches a stop position (1013, and Figure 4, A).

          The sequence window is now incremented in length by one base position and reset to a new start position (1014). The process of evaluation and stepping is again performed until a new stop position is reached (1001-1013). The process of stepping sequence window

30   length and start position is continued until a maximum window length is attained (1015). This process methodically examines the entire examination region (1016).

          Because DNA is double stranded, and the characteristics of the two strands can differ, this process can be repeated for the opposite strand (1016 and Figure 4, C & D). It is also repeated for all of the oligonucleotide functions that are required by the specified model

35   (1017, compare Figures 4, A & B to C & D). This process can also be repeated for more than one molecular biology model per method and one or more methods (1018). It can also

5   be repeated for more than one genetic variation. Using a graphical user interface integrated into the software than carries out the invention, a user navigates between fields and screens and may repeat the process for another target nucleic acid sequence, another target feature of a target nucleic acid sequence, another biochemical method with the same or different target nucleic acid sequence, or any combinations thereof (see Figure 11).

10   If batch processing is to be performed without user intervention, selection criteria typically must be specified so that typically a single oligonucleotide is chosen for each oligonucleotide function otherwise a user may opt to terminate the program (1019).

Examination Region

15   Each molecular biology method will have associated with it one or more Examination Regions, *e.g.*, a segment of nucleic acid adjacent or flanking on either side to the polymorphic region. There is typically one examination region for each oligonucleotide function (*e.g.*, as a primer, hybridization probe, or antisense nucleotide, etc.). Typically, the order of processing of the oligonucleotide functions, and thus examination regions, is

20   important. Examination regions can be sorted into one or more priority classes whereby the one or more examination regions in a higher priority class are processed before the one or more examination regions of lower priority classes. Within priority class, examination regions may be processed in any order. The examination region associated with a lower-priority oligonucleotide function may depend on the specific oligonucleotides chosen for

25   higher-priority functional classes. The examination region need not be specified explicitly by a model, but may be implicit to the model.

Sequence Window

            The parameters of a specified biochemical model are constraints that must be

30   satisfied. Let the finite number of oligonucleotides that satisfy these constraints be the "solution set" or "results set" (the term is used herein interchangeably). Oligonucleotides in the solution set will range in length from some minimum length (L_{min}) to some maximum length (L_{max}). However, the value of L_{min} and L_{max} are not known before hand.

            Because of this, a sequence window (Figure 3, feature B) is defined that ranges in

35   length from W_{min} to W_{max}, where typically $W_{min} \leq L_{min}$ and $W_{max} \leq L_{max}$. Thus, W_{min} and W_{max} are non-constraining in that any shorter or longer window lengths generate only

5   candidate oligonucleotides that fail to satisfy model constraints. W_{min} may be greater than L_{min} and W_{max} may be less than L_{max} to speed computation time if only non-optimal oligonucleotides of the solution set are not evaluated. Typically, W_{min} and W_{max} are chosen such that qualifying candidates begin appearing at window lengths several positions longer than W_{min} and stop appearing several positions short of W_{max}. Due to fast computational

10 speeds of modern computer microprocessors this excess examination typically is not onerous. A person having ordinary skill in the art would be able to determine desired settings of W_{min} and W_{max} without undue experimentation.

Biochemical Models

15 Model variables can include all variables that are well known in the art to influence performance of oligonucleotide annealing, melting, hybridization, and stability. For a single oligonucleotide, this includes but is not limited to T_m (melting temperature in Celsius), 3' stability, length, GC-content, 5' and/or 3' self complementarity and propensity to form secondary structures with itself, the base at 5' position, number of Gs and Cs at the last five

20 3' positions, avoidance of runs of repeated nucleotides or nucleotide types (*e.g.*, purine or pyrimidine), synthesis chemistry, and secondary structure. For more than one oligonucleotide, variables can include pairwise complementarity. For PCR products, variables can include product size and product T_m.

    T_m is typically a critical constraining variable. Many methods specify a minimum

25 and maximum T_m that defines a narrow range. A number of methods, and preferably the nearest neighbor model, can be used to calculate oligonucleotide T_m. The nearest neighboring model depends on experimentally derived values.

    T_m can be calculated for both perfect matches and mismatches. A perfect match arises when an oligonucleotide has perfect Watson-Crick base pairing with its

30 complementary target. A mismatch is when there are one or more mismatches between an oligonucleotide and its complementary target. Currently, nearest neighbor values are available only for single base mismatches. Estimation of mismatch T_m is especially relevant to allele-specific hybridization probes, where an important selection criterion is to maximize the difference between perfect-match T_m and mismatch T_m (ΔT_m). Allelic discrimination

35 typically improves as ΔT_m increases. Perfect match T_m is typically higher than mismatch T_m.

5   Another important model variable is the chemistry used to synthesis oligonucleotides. Phosphoramadite chemistry is standard. In addition, Applied BioSystems, Inc. manufactures a PropyneT chemistry that is marketed as Turbo TaqMan™. PropyneT is a thymidine derivative that increases the nearest neighbor T_m by 1.0° C for every thymidine in the oligonucleotide. PropyneT improves the performance of allele-specific hybridization

10   probes when the probe has a GC content of about 65% or less because PropyneT probes tend to be shorter and attain the targeted minimum T_m, and shorter probes tend to yield higher predicted ΔT_m. Higher predicted ΔT_m typically yield increased allelic discrimination. Derivatives for especially adenine, but also guanine and cytosine, that increase hybridization strength, would be an important means to further improve allelic discrimination by

15   decreasing allele-specific probe length.

A final important variable well known in the art is secondary structure. Secondary structure refers to the stem and loop shapes that oligonucleotides assume due to self-complementarity. Stems are regions of self-complementarity and loops are regions of non-self complementarity. It is known in the art that that a stem and loop structure that looks

20   like a clamp, as embodied in molecular beacons, enhances allele-specific hybridization.

Constraints Imposed by Genetic Variation

An important consideration in assay design for genetic diversity is that oligonucleotides are constrained primarily by the sequence context of the targeted

25   polymorphism or other target feature. Further, choice of forward and reverse PCR primers can be constrained by the need to minimize PCR product length. These two factors can greatly limit examination regions. A biochemical model that has many variables with stringent parameters may frequently generate an empty solution set.

Further, it is important to realize that simple models that take into account only two

30   or three variables have performed adequately in some labs. The present invention is demonstrated using a four variable model that for the 5' nuclease assay has a success rate of > 90%. Models that take into account additional variables may improve the success rate of assay designs.

For model building, the sequence constraints of targeting genetic variation and the

35   success of simple models suggest a forward selection process. That is, to first investigate one or two additional variables that are likely to have substantial impact on performance.

-29-

5

Process

Experience in designing 5' nuclease assays has led to the definition of a set of tasks and an ordering of these tasks. Further, it has led to the definition of a biochemical model with default parameter values that rarely need to be change. Methodically following the set

10   of ordered tasks and using the default parameters results in a successful assay design >90% of the time.

The present invention simplifies this process by reducing it into software via a user interface wizard. This interface methodically leads users through the defined process, thereby performing for the user repetitive and time-consuming tasks. Typically, the user

15   merely needs to use the mouse to point and click through the series of screens. Only rarely will the user need to change parameter settings or make other kinds of manual interventions.

The graphical user interface presents to the user the oligonucleotides in the solution set, typically sorted by one or more variables in the model. The user then can scroll among these lists so as to select their preferred oligonucleotides. Typically, these will be

20   oligonucleotides at, or near, the top of the list. The user interface allows only one oligonucleotide to be selected for each functional category and oligonucleotides must target the same strand.

Max of the Min Selection Criteria

25   For a given oligonucleotide function, there typically are at least several candidate oligonucleotides within each solution set. Further, the user typically has a solution set from more than one model to choose from. Because of this it is useful to define selection criteria (or ranking parameters). Selection criteria are a convenient guide if the user manually selects an oligonucleotide from a solution set. Selection criteria are a necessity if the user

30   desires assay designs to be generated automatically, as would be the case with batch processing of many genetic variations.

Within solution sets, the selection criteria could be to pick the oligonucleotide that appears at the top of the sorted list, assuming the list is sorted with those oligonucleotides having the most optimal characteristics at the top. To choose among more than one model,

35   the model should be chosen for which the oligonucleotides for each function exhibit the most preferred characteristics.

5   The present invention specifies a selection criteria set for the special case in which
two or more allele-specific hybridization probes are being compared among two or more
models. The strategy is to select for each model the probe with the worst characteristics.
The probe among these that has the best characteristics specifies the model that should be
chosen.

10

Examples

**Oligonucleotide Selection for the Applied Biosciences, Inc. "TaqMan™" Assay and
SNP analysis**

         The preferred embodiment in this example is the 5' nuclease method. In this method
15   there are typically four oligonucleotides: 2 allele-specific hybridization probes, 1 forward
PCR primer and 1 reverse PCR primer. The two allele-specific hybridization probes
constitute one priority class and the forward and reverse PCR primers constitute a second
priority class. The allele-specific hybridization probes have higher priority than the PCR
primers so that the PCR primers do not overlap the allele-specific hybridization probes.

20   The alleles of the genetic variation, for which allele-specific hybridization probes are
needed, typically can be processed in any order. Once allele-specific hybridization probes
are chosen, the forward and reverse PCR primers can typically be processed in any order. In
this model, the examination regions for the forward and reverse PCR primers depend on the
exact oligonucleotides chosen as allele-specific hybridization probes.

25

**Table 1A. Standard and Propyne T Chemistry Models and Default Parameter Values
for the Allele-Specific Hybridization Probes of the 5' Nuclease Method**

	Standard	PropyneT
5' Position	No G	No G
Min T_m	68.0° + GC% factor*	68.0°+GC% factor
Max T_m	72.0°	72.0°
Open Positions	Middle Third	Middle Third
Nearest Neighbor Values	Unified	Unified + 1.0*(#Ts)
T Synthesis Chemistry	Phosphoramadite	5-propyne-2'deoxyuridine
ΔT_m	Calculate	Calculate

30   GC % factor is to increase Min T_m (and Max T_m if necessary) some amount, typically about
0.5° - 1.0° C every 10% above 50%.

5   Turning to Figure 5D, a set of potential oligonucleotides are generated for an entered allelic nucleic acid, when a user hits the next button. This operation initiates the comprehensive examination and evaluation algorithm. The first steps of this process are depicted in Figure 4A-D. These steps address the examination regions of the two allele-specific hybridization probes, which are the first priority.

10   In Figure 4, A, a portion of the examination region of the sense strand for allele C is depicted where the variable base is in upper case and all other nucleotide positions are in small case. Also depicted are all possible oligonucleotides of $W_{min}=15$ for which the variable base lies in the middle third of the oligonucleotide. This later constraint is the default parameter setting of the open positions variable of the 5' nuclease model.

15   Mismatches in the middle of an oligonucleotide tend to disrupt hybridization more than mismatches towards either end.

With reference to the numbering system of Figure 2, A & B, the Open Positions constraint defines positions 6-10 of the sequence window as the Open Positions. In turn, the Open Positions parameter dictates that the first start position in the examination region be

20   position −5. The variable position is paired with the sequence window position 6, which is the first open position. In this location, the sequence window defines the first candidate oligonucleotide to be the sequence tcgccCggactccga, which are positions −5 to 9 of the examination region. This candidate is evaluated with respect to the first 5'nuclease model, and saved or discarded.

25   The sequence window is then incremented one base 5' so as to pair the variable base with the next open position, which is position 7. In this location, the sequence window defines the second candidate oligonucleotide as positions −6 to 8 of the examination region. This candidate is evaluated with respect to the 5' nuclease model and saved or discarded. This process is repeated for open positions 8, 9 and 10. The open positions dictate the stop

30   position in the examination region to be position −9.

The sequence window is now incremented one base to length 16 (not shown), and the process examination and evaluation is repeated until the new stop position is reached. The process of incrementing the sequence window is repeated until the window length is equal to $W_{max}=25$. By this method, the first examination region is comprehensively

35   examined, and all oligonucleotides in the solution set are identified.

5   The comprehensive examination and evaluation described above is repeated for the second examination region corresponding to the second allele-specific hybridization probe (Figures 4 A&C vs B&D). It is also repeated for both strands (Figures 4 A&B versus C&D), and for both biochemical models (not shown).

       The solution sets are listed in Figures 5E-H. As can be seen, the oligonucleotides in

10 each solution set are sorted in descending order for the ΔT_m variable. Other pertinent information, such as T_m, GC%, length, and oligonucleotide sequence, is shown. In the present system, the selection criterion is to maximize ΔT_m. Thus, the results that the user first sees are for the antisense strand and the PropyneT chemistry model. This is because the minimum ΔT_m between allele-1 and allele-2 for this model is 7.5. The other pairwise

15 minimum ΔT_ms are 7.5, 3.9 and 3.8. If the solution sets are deemed inadequate, the user can select to return to the parameters screen, change parameters, and re-run the comprehensive examination and evaluation algorithm.

       Figure 5I depicts the users selections. Before display, the selections are evaluated to confirm that the user has made a selection for each required oligonucleotide and that the

20 oligonucleotides target the same strand. The user has the choice only of reporter and quencher dyes, as all other values are calculated. The default values of the reporter and quencher dyes are shown, and these rarely need to be changed. It is possible to select an unmodified oligonucleotide, that is, to select 'None' for the reporter or quencher.

       Clicking on the Next button of the probe selections screen leads to the primer

25 parameters screen (Figure 5J). Here the user can enter W_{min}, W_{max}, Min-T_m and MaxT_m values. The defaults are shown, and these rarely need to be adjusted.

       The examination region of the forward PCR primer is defined to end one base 5' of the most 5' position of the hybridization probes. The beginning of the examination region is either the most 5' position of the inputted flanking sequence, or position −100 relative to the

30 variable position, whichever is reached first. The latter value is for convenience only and could be larger or smaller. The examination region of the reverse PCR primer is defined in an analogous manner.

       Figure 5K shows the solution set for the forward primer and Figure 5L show the solution set for the reverse primer. The forward primers are displayed 5'→ 3' with the

35 oligonucleotides closest to the variable site at the top of the list, and are aligned with the 5' flanking sequence. The reverse primers are displayed 3'→ 5' also with the oligonucleotides

-33-

5   closest to the variable site at the top of the list, and are aligned with the 3' flanking
sequence.

Primer selections are depicted in Figure 5M. Also depicted are the selected allele-
specific hybridization probes and PCR primers aligned with the nucleotide sequence
flanking the variable position. The user has no design choices here as all values are
10   calculated. Clicking on the Save button leads the user to a screen to enter a name for the
assay and other information (Figure 5N). The default name follows the scheme of naming
the assay based on the name of the gene in which the variation lies, the name of the
variation, and a shorthand name for TaqMan, which is "TM". Since this is the first TaqMan
assay to be designed, it is designated as 1. The final screen allows the user to select names
15   for the PCR primer set and each PCR primer.

A cartoon of the examination regions and their functions for the 5' nuclease method
are depicted in Figure 6A. Figure 6B depicts the 5' nuclease method when minor groove
binder proteins are conjugated to the oligonucleotide. Minor groove binder proteins have
been shown to increase ΔT_m and yield excellent allelic discrimination.

20

Table 1B Nearest Neighbor ΔT_m Analysis of 90,000 SNPs Predicts General
Applicability of 5' Nuclease Propyne T Probes

ΔT_m	ND	1	2	3	4	5	6	7
Regular T	17%	2%	14%	29%	22%	11%	4%	1%
Propyne T	0%	0%	1%	10%	32%	40%	15%	2%

The 5' nuclease method of SNP scoring exploits the difference in melting
25   temperature ΔT_m between a probe hybridized with its perfect match complement and the
same probe hybridized with a mismatch complement. To date, the general applicability of
the 5' nuclease method has been limited due to the hit or miss problem of designing probes
with sufficient ΔT_m to achieve robust allelic discrimination. Thermodynamic values for all
possible single base mismatches are available that allow the nearest neighbor model to be
30   applied to SNPs to predict probe ΔT_m. A computer program that maximizes probe ΔT_m has

-34-

been implemented and applied to 102,900 putative SNPs (NCBI's dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), release May 19 2000). The constraints were probes with perfect match T_m between 68° C and 72° C, that have the variable position in the middle third of the probe, that do not begin with a G, and have length between 15 bp and 35 bp. Where No Data indicates no qualifying probe set, the table shows that 67% of SNPs yielded predicted ΔT_ms greater than 3.0° C using standard phosphoramadite chemistry. Three degrees is a rule-of-thumb threshold over which probes often yield discrimination. Replacement of the probe T bases with 5-propyne-2'-deoxyuridine, a T derivative, is predicted to increase T_m 1.0° C for every T in the probe. Propyne T probes have been used successfully in AT rich regions. Applying PropyneT probes for SNPs with flanking GC%<60.0 shortened the average probe from 25.0 bp to 21.5 bp, and increased the average ΔT_m 31% from 3.9° C to 5.1° C. These results suggest that probe sets with maximal ΔT_m can be predicted, and SNPs with ΔT_m less than 3.0° C can be avoided. Further, Propyne T probes predict success of the 5' nuclease method for all but a small fraction of SNPs.

Additional Embodiments

In addition to the 5' nuclease method, the present invention can be applied to most laboratory methods for scoring genetic diversity (Figures 6C-F). These include, but are not limited to, the anchor method, the Invader method, the single base extension method, pyrosequencing, ligation methods, and the DASH method. Applicable to each of these laboratory methods is the generalized method and process of defining examination regions for each oligonucleotide function, identifying the solution set by performing a comprehensive examination and evaluation of the examination regions, examining both strands, comparing solution sets from one or more biochemical models, and picking a single oligonucleotide for each function based on selection criteria.

Implied also as an additional embodiment are different kinds of genetic variations. The example and coded software addresses nucleotide substitutions. Also known as Single Nucleotide Polymorphisms, these are the substitution of one base for another. The present invention can also be employed for insertions and deletions, where one or more nucleotides are deleted or inserted in comparison to a reference sequence.

The application of the algorithm to insertions and deletions should be obvious to a practitioner of the art. Briefly, for allele-specific hybridization type methods, one

-35-

5   hybridization probe is designed to the junction where the sequence is deleted and one hybridization probe is designed to the inserted sequence. For single base extension, the sequencing primer ends at the last base before the insertion.

As described, the invention evaluates each candidate oligonucleotide in isolation. However, the combination of oligonucleotides, one for each function, can impact assay

10   performance. This is because there is an interaction among the oligonucleotides. Another embodiment incorporates a biochemical model in which candidate oligonucleotides for one function are evaluated with respect to candidate oligonucleotides for all other functions. Briefly, in the Overview of Process Flow Chart, Figure 9, the box 'Apply Selection Criteria' would be enhanced to evaluate oligonucleotides among functional categories.

15   The following are the evaluation parameters and constraint values used by the software of the invention in a further embodiment. In this embodiment the biochemical method includes minor groove binding (MGB).

Table 2A: Probes – Gene Expression and DNA Detection

Length	11-18
T_m	69±1
Max Secondary Structure T_m	35
Concentration	0.25 µM
MaxMonoNucRuns	4
Last5Bases	at least 3 Gs or Cs
Max AT run outside MGB	3
GC%	about 20% to about 80%

20   **Table 2B: Probes – SNP Discrimination**

Length	11-18
T_m	66±2
Max Secondary Structure T_m	35
Concentration:	0.25 µM
MaxMonoNucRuns	4
Last5Bases	at least 3 Gs or Cs
Max AT run outside MGB	3
GC%	about 20 to about 80

Table 2C: Primers

T_m:	68 ± 1
Max Secondary Structure T_m:	About 35 Celsius
Concentration: Left: 0.1 µM; Right:	2.0 µM
MaxMonoNucRuns:	3
Avoid	A at 3' end
GC%:	About 40 to about 70
10 bp buffer from probe if on same strand	
Amplicon	
Max GC%	About 65
Max Length	About 170

5

Table 2D: Technology Platform Integration (Epoch supplies software licenses and hardware)

Browser Compatibility	Microsoft Internet Explorer V. 4.5; Netscape Version 4.7
Web Pages	Microsoft Application Server Pages Version 3.0
Web Server	Microsoft IIS V 4.0
Database	Oracle V. 8.7.1
Analytics	Compiled DLL; Microsoft Component Object Model V
Operating System	Windows NT V

10

Relational Database

The processing of tens of thousands of assay designs requires storing targeted variations and their flanking sequence in a database for electronic retrieval by the design software. Further, a database is needed to temporarily store and retrieve the oligonucleotides

15 generated for each model, and to permanently store and retrieve the final oligonucleotides chosen for each design (Figure 8).

A practitioner of the art will appreciate the power provided by the simple table layout to accommodate any genetic variation, any number of designs for a specific variation, and any number of oligonucleotides of stated function for each design.

20 Because typically a team of scientists conducts one or more projects, there is a Group table that records the members of the group and a Project table that can have one or more projects. Each project comprises one or more genes, and each gene typically has one or more variations. Each variation may typically have zero or more designs (Assays) for one or more methods. If an assay requires PCR, it will have a forward and reverse PCR primer that

25 defines the PCR product, which could also be known as a Sequence Tag Site, although here

5   no mapping information is implied. The assay will have one or more oligonucleotides, each with a stated function.

The key and foreign key fields are specified by the label key and foreign key (FK) respectively. The names of the fields typically provide enough information to a practitioner of the art to understand their meaning. Because of this self-description, the field names

10   represent also the data dictionary.

The data tables can be populated primarily electronically. That is, genes and variable sites data can be downloaded from data warehouses. The above-described algorithm can generate Assays, PCR products and Oligos data. In the event that manual data entry is needed, also provided are a large number of Look Up Tables (LUTs), that allow users to

15   select values from a list rather than having to type them in repeatedly.

A key feature of the tables, reflecting the present invention, is the one-to-many relationship between the Variation table and the Assay table. This allows more than one method to be applied to a particular variation, if not in the laboratory, at least at the design stage. The MethodName field in the Assay table contains this information.

20   Another key feature of the tables, reflecting the present invention, is the one-to-many relationship between the Assay table and the Oligos table. It is this relationship that captures the essence of the generality of the invention. In particular, the OligoUse field of the Oligonucleotide table contains the function name pertaining to the function that the oligonucleotide performs for a given method. These names include Forward PCR Primer,

25   Reverse PCR Primer, Allele-Specific Probe, Anchor Probe, Invader Probe, etc. (see Figure 8).

Also important to the Oligos table are the 5' modification fields and 3' modification fields. These fields contain data pertaining to modifications made to the oligonucleotide during or after synthesis, such as conjugating a fluorescent dye, a minor groove binder

30   protein, a mass tag, or a generic sequence for signal amplification or detection.

Substantially Similar Function

A number of variations on this invention would yield identical results. For example, a described comprehensive examination method has the sequence window movement loop

35   inside the sequence window length loop. A comparable strategy is to have the sequence window length loop inside the sequence window movement loop. Similarly, the invention is

-38-

5   described as starting with a minimum sequence window length and incrementing towards a

maximum sequence window length. Starting at a maximum length and incrementing down

to a minimum length is comparable. All such variations are encompassed in within the

invention may employ biochemical methods as the outer loop, function as the middle loop

and sequence strand as the inner loop. This means that the examination regions are

10   separately examined for each model. However, it may be the case that two or more models

share the same functions and examination regions. In such a situation it would be

computationally faster to perform the comprehensive examination once and evaluate each

candidate oligonucleotide multiple times, once for each model.

A feature of the tables that would be evident to a person practiced in the art is that

15   many of the fields are annotations. Some of these fields may be eliminated to accommodate

specific purposes, and many other fields included, such as all GenBank fields, as data and

features are accumulated.

Operability with DNA Synthesizers

20   Another embodiment of the current invention is to link the underlying algorithm and

computer system to a DNA synthesizer such that the reagent oligonucleotide sequences

generated can be inputted into the synthesizer to make the desired probes or primers and

further automate the process. DNA synthesizer software generally accepts sequence

information in the form of text or ASCII, most accept sequence in a plurality of computer

25   encoded formats. A person having ordinary skill in the art recognizes that computerized text

information can be readily converted from one format to another through text filters or word

processor embedded conversion software. DNA synthesizers can be made operatively

linked to a computer system directly or remotely by any means known in the art (*e.g.*, serial

or parallel connectivity, SCSI, SCSI-2, Cray Cabling, Universal Serial Bus, coax cabling,

30   fiber optic cabling, all forms of computer network connectivity, etc...). Commercially

available DNA synthesizers include but are not limited to MerMade II™ (BioAutomation),

Perkin Elmer/Applied Biosciences, Inc. Model 394/5 DNA Synthesizer, Labtronix DNA

synthezier, or ASM-700 DNA synthesizer (BIOSSET).

The embodiments illustrated and discussed in this specification are intended only to

35   teach those skilled in the art the best way known to the inventors to make and use the

invention. Nothing in this specification should be considered as limiting the scope of the

-39-

5   present invention. The above-described embodiments of the invention may be modified or varied, and elements added or omitted, without departing from the invention, as appreciated by those skilled in the art in light of the above teachings. It is therefore to be understood that, within the scope of the claims and their equivalents, the invention may be practiced otherwise than as specifically described.

10

The following and any foregoing references and publications are herein incorporated by reference in their entirety:

Allawi H.T. & SantaLucia Jr, J. (1998) Nearest-Neighbor Thermodynamic Parameters for Internal G-A Mismatches in DNA. Biochemistry (37) 2170-2179.

15

Allawi H.T. & SantaLucia Jr, J. (1998) Thermodynamics of Internal C-T Mismatches in DNA. Nucleic Acids Research 26(11):2694-2701.

Allawi H.T. & SantaLucia Jr. J. (1998) Nearest-Neighbor Thermodynamics of Internal A-C
20   Mismatches in DNA: Sequence Dependence and pH Effects. Biochemistry (37) 9435-9444.

Allawi H.T. & SantaLucia, Jr. J. (1997) Thermodynamics and NMR of Internal G-T Mismatches in DNA. Biochemistry 36:10581-10594.

25

Balaban DJ et al. (1999) System for Providing a Polymorphism Database. WO 99/05324.

Landegren U, Nilsson M, & Kwok P-Y (1998) Reading Bits of Genetic Information: Methods for Single-Nucleotide Polymorphism Analysis. Genome Research 8:769-776.

30

Livak KJ (1999) Allelic discrimination using the fluorescent probes and the 5' nuclease assay. Genetic Analysis: Biomolecular Engineering (14) 143-149.

Morin P.A., Saiz R. & Monjazeb A (1999) High-throughput Single Nucleotide
35   Polymorphism Genotyping by Fluorescent 5' Exonuclease Assay. BioTechniques 27:538-552.

Peyret, N., Seneviratne, P.A., Allawi, H.T. & SantaLucia, Jr. J (1999) Nearest Neighbor Thermodynamics and NMR of DNA Sequences with Internal A-A, C-C, G-G and T-T
40   Mismatches. Biochemistry (38) 3468-3477.

Sabatini et al. (1999) Database and Systems for Determining, Storing and Displaying Gene Locus Information. US Patent #5,970,500.

45   SantaLucia, Jr. J. (1998) A Unified View of Polymer, Dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. (95) 1460-1465.

5   Yuan, C.C., Peterson R.J., Wang C-D, Goodsaid, F, Waters, D.J. (2000) 5' Nuclease Assays
     for the Loci CCR5-+/D32, CCR2-V64I, and SDF1-G801A Related to Pathogenesis of
     AIDS. Clinical Chemistry 46(1):24-30.

<u>**WHAT IS CLAIMED IS:**</u>

1. A method for determining an optimal reagent oligonucleotide sequence for use in a biochemical method for evaluating a target nucleic acid sequence having a target feature, the method comprising the steps of:

    defining a set of exclusion values and/or ranking values specific to the biochemical method,

    defining a sequence window adjacent to the target feature,

    generating candidate reagent oligonucleotide sequences complementary to one or both of the sense and antisense strands of the target nucleic acid sequence within the sequence window, the reagent oligonucleotide sequences having a length less than or equal to the sequence window,

    evaluating the candidate reagent oligonucleotide sequences against the exclusion and/or ranking parameters,

    selecting at least one optimal reagent oligonucleotide sequence for the selected biochemical method as applied to the target nucleic acid sequence.

2. The method of claim 1, further comprising:

    selecting the target nucleic acid sequence, the sequence having at least one target feature, and

    selecting the biochemical method for evaluating the target nucleic acid sequence.

3. The method of claim 1, comprising evaluating candidate oligonucleotide sequences against both exclusion and ranking values.

4. The method of claim 1, comprising generating more than one type of candidate sequences and evaluating the more than one type of sequences, the evaluating comprising evaluating the candidate sequences against compatibility values, and further comprising selecting an optimal set of reagent nucleotides comprising more than one oligonucleotide sequence.

5. The method of claim 1, wherein the generating step comprises stepping the sequence window along an examination region of the target sequence.

6. The method of claim 1, wherein the generating step comprises selecting all possible reagent oligonucleotide sequences within the sequence window.

5   7. The method of claim 1, wherein the sequence window has a minimum length sufficient to generate candidate reagent oligonucleotides specific for the target nucleic acid sequence,

8. The method of claim 1, comprising:

discarding a generated reagent oligonucleotide sequence if it falls outside the range
10   of exclusion constraint values, and saving a generated reagent oligonucleotide sequence if it falls within the exclusion constraint values, and ranking the saved reagent oligonucleotide sequences based on comparison to one or more ranking values thereby providing several reagent oligonucleotide sequences, and selecting from the ranked reagent oligonucleotide sequences optimal reagent oligonucleotide sequences for the
15   selected biochemical method as applied to the target nucleic acid sequence.

9. The method of claim 1 further comprising cataloguing the target nucleic acid sequences together with their corresponding optimal set of reagent oligonucleotide sequences and the corresponding biochemical method in a computer database.
20

10. The method of claim 1 wherein at least one generated reagent oligonucleotide sequence comprises a sequence complementary to the target feature.

11. The method of claim 1 comprising selecting a plurality of biochemical methods each
25   having evaluation parameters and having exclusion, ranking, and/or compatibility values, and further comprising generating and evaluating candidate sequences for one of the biochemical methods.

12. The method of claim 1 comprising selecting at least one reagent oligonucleotide
30   sequence from the group consisting of a forward and reverse primer pair and one or more probes.

13. The method of claim 1 wherein the biochemical method is selected from the group consisting of: the polymerase chain reaction, propyneT chemistry, phosphoramadite
35   chemistry, reverse transcriptase-polymerase chain, ReactionNucleotide™ Sequencing, the anchor method, the Invader method, the single base extension method, cycle

-43-

5   sequencing, cyclical polymerase chain reaction, pyrosequencing, ligation, fluorescent in
     situ hybridization, allele-specific oligonucleotide hybridization (ASOH), dynamic allele-
     specific hybridization (DASH), antisense oligonucleotide chemistry, nucleic acid hybrid
     chemistry, and DNA/RNA repair.

10   14. The method of claim 1 wherein the evaluation parameters are selected from the group
     consisting of: OligoTm, BufferMg++, BufferK+, DivalentMultiplier, Oligonucleotide
     Concentration, AmpliconTm, Amplicon Length, AmpliconGC, OligoLength, OligoGC,
     OligoMonoNucRunLength, Oligo5EndLinker, Oligo5EndModification, Oligo5EndTail,
     Oligo5EndAllowedBases, Oligo5EndLeftPosition, Oligo5EndRightPosition,
15   Oligo3EndLinker, Oligo3EndModification, Oligo3EndTail, Oligo3EndAllowedBases,
     Oligo3EndLeftPosition, Oligo3EndRightPosition, Oligo3EndAnalysisLength,
     Oligo3EndGPlusC, Oligo3EndDeltaG, Oligo3EndGCClampLength, OligoAlignMatch,
     OligoAlignMisMatch, OligoAlignBulgePenalty, OligoAlignMaxBulgeAllowed,
     OligoHairPinMinStemLength, OligoHairPinMinLoopLength, OligoHairPinScore,
20   OligoHairPinDeltaG, OligoHairPinTm, OligoPairAlignScore, OligoPairAlignDeltaG,
     OligoPairAlignTm, Oligo3EndPairAlignScore, Oligo3EndPairAlignDeltaG,
     Oligo3EndPairAlignTm, SequenceFeature5EndBufferLength,
     SequenceFeature3EndBufferLength, SequenceFeaturePosition5End,
     SequenceFeaturePosition3End, and Library SequenceAlignScore.

25

     15. The method of claim 1 wherein the reagent oligonucleotides have a length between 10
     and 50 base pairs.

     16. The method of claim 1 wherein the reagent oligonucleotides have a T_m of about 50 to
30   about 80 degrees Celsius.

     17. The method of claim 1 wherein the reagent oligonucleotides have a guanine-cytosine
     content between about 10% to about 90%.

35   18. The method of claim 1 wherein the reagent oligonucleotides have a maximum secondary
     structure T_m of about 35 degrees Celsius.

-44-

5

19. The method of claim 1 wherein an exclusion value is that the reagent oligonucleotide sequence does not contain an adenine as the ultimate 3' base.

20. The method of claim 1 wherein an exclusion value is that the reagent oligonucleotide
10   sequence does not contain a guanine as the ultimate 5' base.

21. The method of claim 1 wherein the target nucleic acid sequence is entered by selection from a database.

15   22. The method of claim 1 wherein the target nucleic acid sequence comprises more than one target feature.

23. The method of claim 22 wherein the method steps are repeated until optimal reagent oligonucleotide sequences have been generated for all target features and their
20   corresponding biochemical methods.

24. The method of claim 1 wherein the target feature of the target nucleic acid sequence is selected from the group consisting of: a single nucleotide polymorphism, a multimeric subsequence of the target nucleic acid sequence, a cloning sequence, the entire target
25   nucleic acid sequence, a codon, an exon, an intron, a telomere, a viral sequence, a transposon, a noncoding region, a promoter, an enhancer sequence, an expressed sequence tag, and a sequence tagged site.

25. The method of claim 1 wherein the reagent oligonucleotide sequence is selected from the
30   group consisting of: an amplification primer, a sequencing primer, a sequence specific hybridization probe, an anchor probe, an invader probe, a reporter probe, and an antisense oligonucleotide.

26. A computer readable data storage medium storing a computer readable program code
35   means for causing a computer to perform the steps of the method in claim 1.

5   27. A computer readable data storage medium storing a computer readable program code
means for causing a computer to perform the steps of the method in claim 9.

28. A computer system comprising:

a database storing a plurality of target nucleic acid sequences having at least

10   one target feature;

a graphical user interface including:

means for selecting a said target nucleic acid sequence from said database;

means for selecting a biochemical method for evaluating the selected target

nucleic acid sequence ; and

15   means for displaying reagent oligonucleotide sequences that satisfy the exclusion

and ranking parameters of said biochemical method; and

a computer-readable data storage medium comprising program code means for

defining a set of exclusion values and/or ranking values specific to the biochemical

method, defining a sequence window adjacent to the target feature, generating

20   candidate reagent oligonucleotide sequences complementary to one or both of the

sense and antisense strands of the target nucleic acid sequence within the sequence

window, the reagent oligonucleotide sequences having a length less than or equal to

the sequence window, evaluating the candidate reagent oligonucleotide sequences

against the exclusion and/or ranking parameters, selecting at least one optimal

25   reagent oligonucleotide sequence for the selected biochemical method as applied to

the target nucleic acid sequence.

29. The computer system of claim 28, wherein the graphical user interface further comprises

means for selecting from the displayed reagent oligonucleotide sequence.

30

30. A process of manufacturing reagent oligonucleotides comprising using reagent

oligonucleotide sequences from the computer system of claims 28 or 29 in a nucleic acid

synthesizer to produce the selected reagent oligonucleotides.

5   31. A kit of a predetermined number of reagent oligonucleotides optimized for a biochemical method used in evaluating a target nucleic acid sequence, the reagent oligonucleotides made by the process of claims 30.

32. The kit of claim 31 wherein the predetermined number is one or more.

10

33. A method of ordering a kit of reagent oligonucleotides comprising:

accessing a computer system of claims 28 or 29,

inputting the desired target nucleic acid sequence, wherein the target nucleic acid sequence has at least one target feature,

15   generating a set of reagent oligonucleotide sequences useful in one or more biochemical methods used in evaluating said target nucleic acid sequence,

selecting desired sequences from the list of generated reagent oligonucleotide sequences,

synthesizing a kit of reagent oligonucleotides based on the selected reagent

20   oligonucleotide sequences,

shipping said kit of reagent oligonucleotides.

34. The method of claim 33 further comprising entering payment information into said computer system.

TGGTCATCGTGGCCATCGCC[C/T]GGACTCCGAGACTCCAGACC

FIG. 1A

tggtcatcgtggccatcgccYggactccgagactccagacc

FIG. 1B

tggtcatcgtggccatcgccCggactccgagactccagacc
tggtcatcgtggccatcgccTggactccgagactccagacc

FIG. 1C

21TtggtcatcgtggccatcgccCggactccgagactccagacc

FIG. 1D

5'- g c c a t c g c c T g g a c t c c c g a -3'
    -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9

FIG. 2A

5'- g c c a t c g c c T g g a c t -3'
    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

FIG. 2B
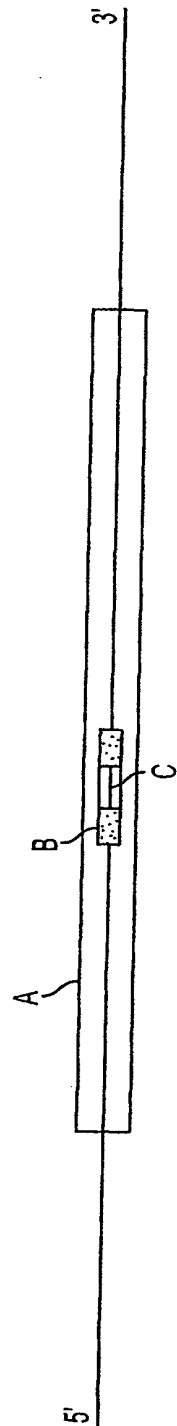
FIG. 3

```
5'-tggtcatcgtggccatcgccCggactccgagactccagacc-3'
            tcgccCggactccga
           atcgccCggactccg
          catcgccCggactcc
         ccatcgccCggactc
        gccatcgccCggact
```

FIG. 4A

```
5'-tggtcatcgtggccatcgccTggactccgagactccagacc-3'
            tcgccTggactccga
           atcgccTggactccg
          catcgccTggactcc
         ccatcgccTggactc
        gccatcgccTggact
```

FIG. 4B

```
5'-ggtctggagtctcggagtccGggcgatggccacgatgacca-3'
               agtccGggcgatggc
              gagtccGggcgatgg
             ggagtccGggcgatg
            cggagtccGggcgat
           tcggagtccGggcga
```

FIG. 4C

```
5'-ggtctggagtctcggagtccAggcgatggccacgatgacca-3'
               agtccAggcgatggc
              gagtccAggcgatgg
             ggagtccAggcgatg
            cggagtccAggcgat
           tcggagtccAggcga
```

FIG. 4D
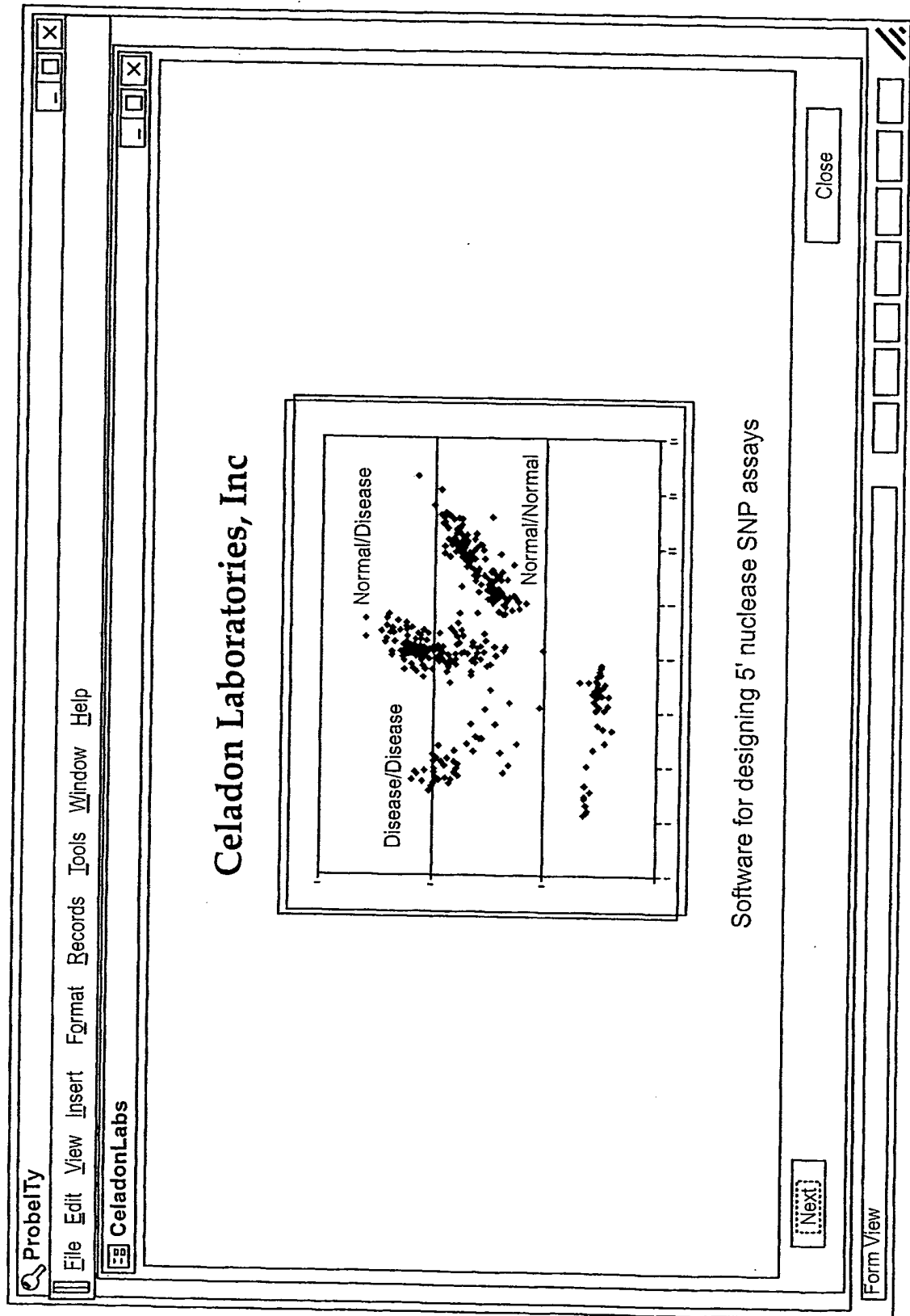
FIG. 5A

ProbelTy

File Edit View Insert Format Records Tools Window Help

Genes

Customer Name: Celadon

Gene Symbol	Gene Name	Species
ABCR		HUMAN
APOE		HUMAN
ARNT		HUMAN
BRL1		HUMAN
CCR2		HUMAN
CYP2E1		HUMAN
CYP3A4		HUMAN
DCSIGNP		HUMAN
IGF1		HUMAN
IL10		HUMAN
IL1A		HUMAN
IL1B		HUMAN
IL2		HUMAN

Add New Gene

Cancel   < Back   Next >   Finish

Form View

FIG. 5B

FIG. 5C

FIG. 5D

FIG. 5E

FIG. 5F

FIG. 5G

FIG. 5H

FIG. 5I

FIG. 5J

ProbeITy

File Edit View Insert Format Records Tools Window Help

CeladonLabs

| Primer Parameters | Forward | Reverse |

Tm GC% Length

Tm	GC%	Length	5'-Forward-3'
59.2	56	16	GTGCATAGCCTCTCCT
59.4	53	17	GTGCATAGCCTCTCCTA
58.5	60	15	GGAAGTGGAGGTCGT
59.3	56	16	TTGGGAAGTGGAGGTC
59.3	56	16	CTTGGGAAGTGGAGGT
59.3	56	16	ACCTTGGGAAGTGGAG
59.3	56	16	GACCTTGGGAAGTGGA
59.3	56	16	AGACCTTGGGAAGTGG
59.3	56	16	GGAGACCTTGGGAAGT
58.9	56	16	TGGAGACCTTGGGAAG
58.9	56	16	CTGGAGACCTTGGGAA
58.9	56	16	CCTGGAGACCTTGGG
58.9	67	15	GCACCTGGAGACCTT
58.8	60	15	CTCCAGCACCTGGAG
59.4	67	15	

Next   Design Primers   Close

Form View

FIG. 5K

FIG.5L

FIG. 5M

FIG. 5N

FIG. 50

FIG. 6A

FIG. 6B

FIG. 6C

FIG. 6D

FIG. 6E

FIG. 6F

FIG. 7P

FIG. 7Q

FIG. 7R

Celadon Laboratories, Inc. | ProbeITy | Analyze Sequence - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

```
321 GTTATGTATA TCCAGTATCT GCCCCATGGT AG[A/G]TTCCTTT
361 TTTTTGTTGT TGGAGTTTTG CTCTTGTTGC CCAGGTTGGA
401 GTGCAATGGT GCAATCTTGG CTCACTGCAA CCTCAGCCTG
441 CTGGGTTCAA GCAATTCTCC TGCCTCAGCC TCCCGAGTAG
481 CTGGGATTAA GGCACGTGCC ACCACACCCG GCTAATTTTT
521 GTATTTTCAG TAGAGATGGG GGTTTCACCA TATTGGCCGG
561 GCTGGTCTCG AGCTCCTGAC CTCAGGTGAT CGGCCCATCT
601 CAGCCTCCCA AGTGCTGGGA TTACAGGTGT GAGCCACTGT
641 GCCTGGCCAG ATG
```

Sequence Stats:

	5'	3'	length
Upstream:	1	352	352
SNP:	353	353	1
Downstream:	354	653	300

Search Regions:

	5'	3'	length
Left:	236	342	107
SNP:	329	377	49
Right:	364	471	108

Next>>

<<Back

|<Start Over

Contact Us | Privacy | © Celadon Laboratories, Inc. 2001

Powered by Appx Digital Systems Inc.

Internet

Done

FIG. 7S

FIG. 7T
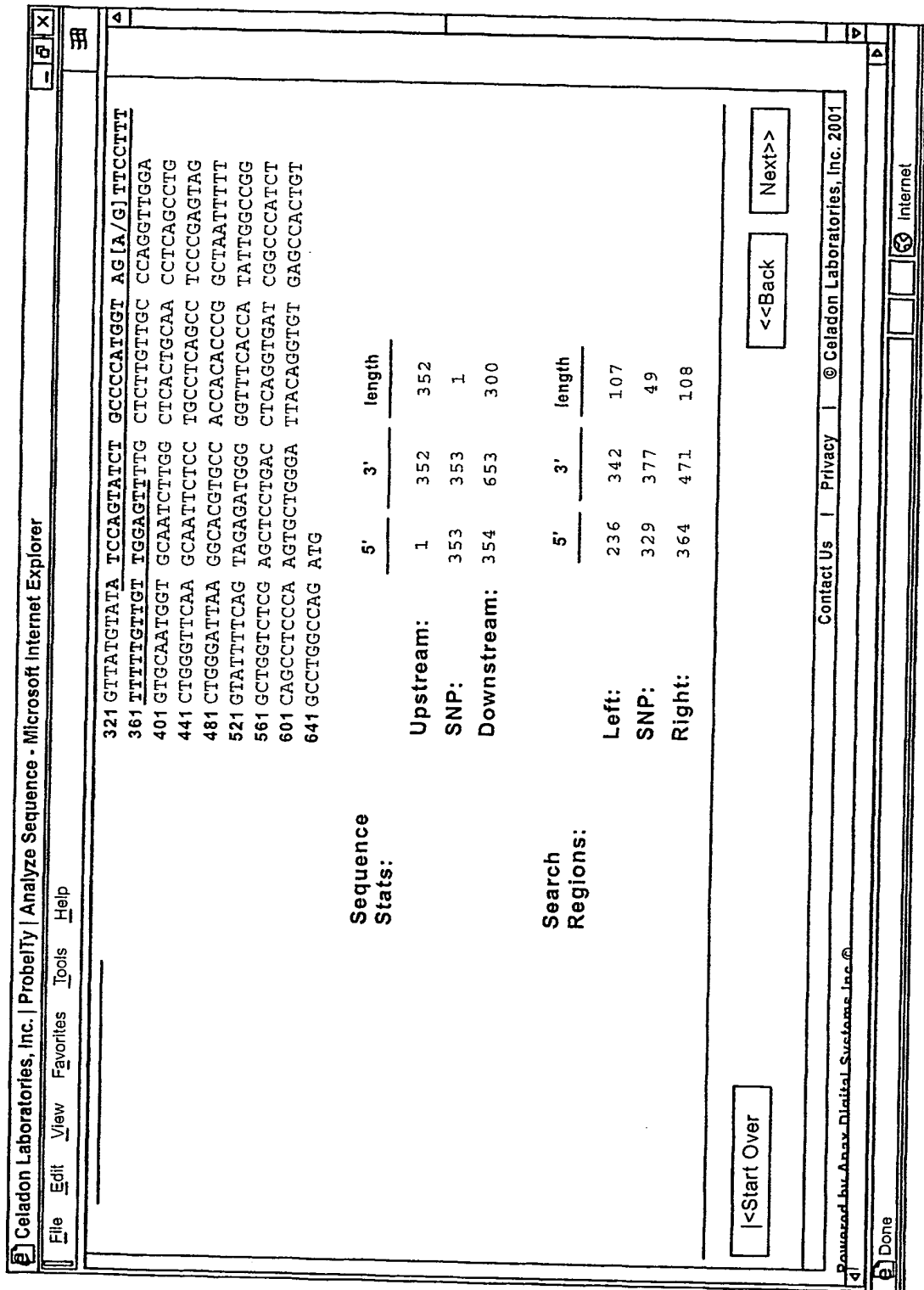
Celadon Laboratories, Inc. | ProbeITy | Edit Parameters - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

ProbeITy™ 3.0

Step 3a of 7

Step 3a: The parameters below are default parameters used in the identified parameter set. To modify any of these parameters, enter the desired values in the appropriate fields.

To proceed, click "NEXT>>".

1. Input Sequence
2. Analyze Sequence
3. Select Parameters

3a. Edit Parameters

4. Select Eclipse Probes
5. Select Primers
6. Select Assays
7. Check Out

click here for more news about
SNPs, Genotyping and Gene Expression

| Review Sequence Details |

Target Parameters

Parameter	Value
SNP 5 Buffer	10
SNP 3 Buffer	10

Oligo Set Parameters

Parameter	Amplicon Set	Probe Set
Amplicon Min Len	50	N/A
Amplicon Max Len	150	N/A
MgCl2 Conc	2.5	2.5
KCl Conc	50	50
Divalent Multiplier	100	100
Max Pair Align Score	8	8

Oligo Parameters

Done   Internet

FIG. 7U

FIG. 7V

FIG. 8A

FIG. 8B

FIG. 8C

(TO FIG. 8C)

(TO FIG. 8C)

(TO FIG. 8A)

PARAM_SET	
PK	PARAMSET_ID
	PARAMSET_NAME

ORDER_OLIGO	
PK, FK2	ORDER_ID
PK, FK1	OLIGO_ID

(TO FIG. 8E)

RESERVED_LOGINS	
PK	RESERVED_LOGIN_NAME

(FROM FIG. 8C)

OLIGO	
PK	OLIGO_ID
FK3,I3	DESIGNRUN_ID
I6	TARGET_ID
FK4,I4	OLIGOTYPE_ID
FK1,I1	BIOSEQ_ID
FK5,I5	STRAND_ID
FK2,I2	CHEMISTRY_ID
	OLIGO_POSITION_5_END
	OLIGO_PM_TM
	OLIGO_PM_TM_2
	OLIGO_MM_TM
	OLIGO_DELTA_TM
	OLIGO_GC
	OLIGO_BEST_LOCAL_ALIGN
	OLIGO_APPROVED
	OLIGO_POSITION_3_END
	OLIGO_SCORE
	DYE_ID_5_END
	DYE_ID_3_END
	OLIGO_FORORDER

(TO FIG. 8E)

(FROM FIG. 8E) (TO FIG. 8E)

STRAND	
PK	STRAND_ID
	STRAND_NAME
	STRAND_DESC

NEAREST NEIGHBOR	
PK	NN_ID
	CHEMISTRY_ID_OLIGO
	CHEMISTRY_ID_TARGET
	LABORIGIN_ID
	NN_SEQ
	OLIGO_SEQ
	TARGET_SEQ
	NN_DELTA_H
	NN_DELTA_S
	NN_DELTA_G
	NN_LITERATURE_REFERENCE

(TO FIG. 8C)

PARAM_SET_PARAM_OLIGOSET	
PK,FK3	PARAMSET_ID
PK,FK2	PARAM_ID
PK,FK1	OLIGOSETTYPE_ID
	PARAM_SET_VALUE

(FROM FIG. 8C)

(FROM FIG. 8E) (FROM FIG. 8E)

(FROM FIG. 8E)

PARAM_LEVEL	
PK	PARAMLEVEL_ID
	PARAMLEVEL_NAME

ASSAY	
PK	ASSAY_ID
FK1,I1	DESIGNRUN_ID
FK2,I2	OLIGOSET_ID_1
FK3,I3	OLIGOSET_ID_2
FK4,I4	OLIGOSET_ID_3
FK5,I5	OLIGOSET_ID_4
	ASSAY_NAME
	ASSAY_APPROVED
	ASSAY_SCORE

(TO FIG. 8F)

(FROM FIG. 8F)

PARAMETER	
PK	PARAM_ID
FK1,I1	PARAMLEVEL_ID
	PARAM_NAME
	PARAM_SORTORDER
	PARAM_METHOD_SPECIFIC
	PARAM_DEFAULT_VALUE
	PARAM_DEFAULT_DESC

FIG. 8D

(FROM FIG. 8F)

(TO FIG. 8G) (FROM FIG. 8G) (FROM FIG. 8G) (FROM FIG. 8G) (TO FIG. 8G) (TO FIG. 8G) (TO FIG. 8G)

SUBSTITUTE SHEET (RULE 26)

FIG. 8E

FIG. 8F

FIG. 8G

PARAM_TARGET

PK,FK1	DESIGNRUN_ID
PK,FK2, I1	PARAM_ID
	PARAMTARGET_VALUE

PARAM_SET_PARAM_OLIGO

PK,FK3	PARAMSET_ID
PK,FK2	PARAM_ID
PK,FK1	OLIGOTYPE_ID
	PARAM_SET_VALUE

OLIGO_SET

PK	OLIGOSET_ID
FK2,I2	OLIGO_ID_1
FK3,I3	OLIGO_ID_2
FK4,I4	OLIGO_ID_3
FK5,I5	OLIGO_ID_4
	OLIGOSET_NAME
	OLIGOSET_APPROVED
	OLIGOSETTYPE_ID
FK1	OLIGOSET_SCORE

METHOD_PARAMSET

PK,FK1,I1	METHOD_ID
PK,FK2,I2	PARAMSET_ID
	METHPARAMSET_DEFAULT

PARAM_SET_PARAM

PK,FK2	PARAMSET_ID
PK,FK1,I1	PARAM_ID
	PARAM_SET_VALUE

(FROM FIG. 8F) (TO FIG. 8F)

(FROM FIG. 8F) (TO FIG. 8F)

(FROM FIG. 8D)

(FROM FIG. 8D)

(TO FIG. 8D)

(TO FIG. 8D)

(FROM FIG. 8D) (TO FIG. 8D)

SUBSTITUTE SHEET (RULE 26)

FIG. 9

FIG. 10

Figure 11

[Continued on next page]

(54) Title: COMPUTER SYSTEM FOR DESIGNING OLIGONUCLEOTIDES USED IN BIOCHEMICAL METHODS

(57) Abstract: A method (1000)for determining one or a set of optimal reagent oligonucleotide sequences used in a biochemical method for evaluating a target nucleic acid sequence having a target feature by defining a set of exclusion values or ranking values specific to the biochemical method (1002, 1005), generating candidate reagent oligonucleotide sequences (1007), evaluating the candidate reagent oligonucleotide sequences against the exclusion or ranking parameters (1008), selecting at least one optimal reagent oligonucleotide sequence for the selected biochemical method as applied to the target nucleic acid sequence (1009).

WO 02/029379 A3

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
15 August 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7)   :G06G 7/48, 7/58
US CL   :703/11
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. :   703/11

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

IEEE, Science Server, EAST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, P	US 6,251,588 B1 (SHANNON et al.) 26 June 2001, figs. 1, 8, col. 18, line 59 to col. 39, line 45. Note col. 83-656.	1-34
Y	US 5,695,940 (DRMANAC et al.) 09 December 1997, fig. 1-2, col. 2, line 51 to col. 4, line 59, col. 11, line 40 to col. 13, line 14, col. 19, line 19 to col. 30, line 42.	1-34
Y	WALTON. S.P. Prediction of Antisense Oligonucleotide Binding Affinity and Activity in Cell Culture. Proc. first Joint BMES/EMBS Conf. October 1999. abstract, page 96.	1-34

☐   Further documents are listed in the continuation of Box C.   ☐   See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)		
		"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 FEBRUARY 2002	29 MAY 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer HUGH JONES
Facsimile No. (703) 305-3230	Telephone No. (703) 305-9704

Form PCT/ISA/210 (second sheet) (July 1998)*